The structure of behavioral variation within a genotype

Zach Werkhoven[1], **Alyssa Bravin**[1], **Kyobi Skutt-Kakaria**[1,2], **Pablo Reimers**[1,3], **Luisa F. Pallares**[4], **Julien Ayroles**[4], **Benjamin de Bivort**[1*]

1 — Center for Brain Science and Department of Organismic and Evolutionary Biology, Cambridge, MA 02138

2 — Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125

3 — Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA 20147

4 — Department of Ecology and Evolutionary Biology and Lewis-Sigler Institute for Integrative Genomics, Princeton University, 08540 Princeton, NJ, USA

* — correspondence: debivort@oeb.harvard.edu

Abstract

Individual animals vary in their behaviors. This is true even when they share the same genotype and were reared in the same environment. Clusters of covarying behaviors constitute behavioral syndromes, and an individual's position along such axes of covariation is a representation of their personality. Despite these conceptual frameworks, the structure of behavioral covariation within a genotype is essentially uncharacterized and its mechanistic origins unknown. Passing hundreds of inbred *Drosophila* individuals through an experimental pipeline that captured hundreds of behavioral measures, we found correlations only between sparse pairs of behaviors. Thus, the space of behavioral variation has many independent dimensions. Manipulating the physiology of the brain, and specific neural populations, altered specific correlations. We also observed that variation in gene expression can predict an individual's position on some behavior axes. This work represents the first steps in understanding the biological mechanisms determining the structure of behavioral variation within a genotype.

Keywords: *Drosophila melanogaster*, high-throughput behavior, individuality, personality, covariation, neural circuits, inbred animals

Introduction

Individuals display idiosyncratic differences in behavior that often persist through time and are robust to situational context. Some persistent individual behavioral traits commonly occur in correlated groups and can therefore be said to covary. Behavioral ethologists have long understood that types of human and animal personalities often fall on multivariate axes of variation. A five dimensional model (Goldberg, 1993) known as The Big Five personality traits is frequently used by psychologists to describe the range of human personality, and a similar model has been used to describe personality in fish (Réale et al., 2007). For example, human propensity for behaviors such as assertiveness, talkativeness, and impulsiveness are collectively described as extraversion and are thought to be anticorrelated with behaviors such as passivity, shyness, and deliberateness, all behaviors associated with introversion (Matthews et al. 2003).

In animal species, correlated suites of behaviors are described as behavioral syndromes. Aggressive behaviors such as fighting over mates or food are frequently correlated with exploratory behaviors such as foraging, and social interaction. Correlated aggressive and exploratory behaviors have been observed in insects (Jeanson and Weidenmüller, 2014), arachnids (Johnson et al. 2007), fish (Huntington, 1976), and birds (Oers et al.

2004). The prevalence of behavioral correlation in so many species suggests that covariation is likely a universal feature of behavior. Although correlated individual differences in behavior are commonly observed, the structure and mechanisms of behavioral covariation are not well understood. Typically, where an individual lands on these behavioral axes is thought to be established by a deterministic confluence of genetic and environmental effects. But there is increasing evidence that substantial individual behavioral variation is rooted in intragenotypic variation (Honegger and de Bivort, 2018; Akhund-Zade et al., 2019). The extent to which intragenotypic behavioral variation is organized into syndromes or axes is essentially uncharacterized.

Substantial variation in specific behavioral measures, even in inbred lines raised in standardized conditions has been observed in several clonal animals, including: geckoes (Sakai, 2018), amazonian mollies (Bierbach et al., 2017), aphids and nematodes (Schuett et al., 2011; Stern et al., 2017). Genetic model systems hold particular promise for the mechanistic dissection of this variation, and intragenotypic variability (IGV) in behavior has been characterized in mice (Freund et al., 2013), zebrafish (Pantoja et al., 2016) and *Drosophila*. In flies, IGV of many behaviors has been studied, including: phototaxis (Kain et al., 2012), locomotor handedness and wing-folding (Buchanan et al., 2015), spontaneous microbehaviors (Kain et al., 2013, Todd et al., 2017), thermal preference (Kain et al., 2015) and object-fixated locomotion (Linneweber et al., 2020). Mechanistic studies of these behavioral phenomena have addressed two major questions: 1) what biological mechanisms underlie the magnitude of behavioral variability (e.g., genetic variation (Ayroles et al., 2015), or neural state variation (Kain et al., 2012; Buchanan et al., 2015)), and 2) what specific differences within individual nervous systems predict individual behavioral biases (Linneweber et al., 2019, Mellert et al., 2016). The mechanistic basis of individuality is an exciting new field, but no study to date has focused on characterizing the large-scale variance-covariance structure of IGV in behavior.

Behavioral correlations within a genotype could arise through a number of biological mechanisms including cell-to-cell variation in gene expression or individual differences in neural circuit wiring or synaptic weights. For example, stochastic variation during developmental critical windows has the potential to impart lasting differences between individuals in the absence of conspicuous genetic or environmental differences. Any such differences affecting nodes common to multiple behaviors in neural or molecular pathways may result in correlated shifts in behavior. Here, we study this directly by focusing on the correlation structure of behavioral variation when genetic and environmental variation are minimized.

This is an important biological question for several reasons. The structure of intragenotypic behavioral variability: 1) will shape the distribution and kinds of personalities that a population of organisms displays, even when they have matched genomes and environments, 2) is the product of stochastic biological outcomes, and its organization reveals how stochasticity drives variation, 3) constrains the evolution of behavior and adaptive phenotypic strategies like bet-hedging (Hopper, 1999), 4) is a relatively uncharacterized component of neural diversity and its manifestations in behavior and disease, and 5) may shed light on how the nervous system orchestrates behavior as a whole. In flies we have a suitable experimental system for directly characterizing this structure, as we can produce large numbers of individuals with nearly identical genomes, reared in the same environment, and collect many behavioral measures per individual. We performed this experiment in wild type inbred flies as well as wild type outbred flies and collections of transgenic lines manipulating neural activity. This approach let us contrast the structure of intragenotypic behavioral variability in animals where the source of variability is, respectively, stochastic fluctuations, genetic differences + stochastic fluctuations, and systematic perturbations of the nervous system + stochastic fluctuations. We found that in all cases, behavioral variation has high dimensionality, i.e., many independent axes of variation. The addition of variation from genetic differences and neural perturbations did not fundamentally alter this qualitative result, suggesting that stochastic fluctuations and genetic differences may structure behavior through common biological mechanisms.

Results

High Dimensional Measurement of Individual Behaviors

The first step in revealing the structure of behavioral variation within a genotype is to devise an experimental pipeline that produces a data set of many (200+) individual flies, with many behavioral measurements each. We first developed a number of behavioral assays, measuring both spontaneous and stimulus-evoked responses of individual flies, which could be implemented in a common experimental platform (Figure 1A; Werkhoven et al., 2019). This instrument features an imaging plane, within which flies moved in arenas of various geometries. Fly position was tracked with digital cameras using diffused infrared illumination invisible to the flies. Visual stimuli were presented to the animals using DLP projectors or LEDs embedded in the arena walls. We implemented six assays in this style, assessing 1) spontaneous walking in circular arenas, 2) preference to rest in brighter or dimmer positions (in an environment of spatially structured illumination), 3) preference to rest in brighter or dimmer light levels (in a fictive, temporally-modulated light environment), 4) optomotor responses to rotating visual stripes, 5) spontaneous left-right decision making in Y-mazes, and 6) phototaxis in Y-mazes, where flies are given a choice of walking toward or away from a lit LED (Figure 1B).

To these assays, we added three more, assessing 7) odor sensitivity in linear chambers (Claridge-Chang et al., 2009; Honegger et al., 2019) in which half of the compartment is filled with an aversive odorant, 8) spontaneous behavior, acquired via high-resolution 100Hz video and suitable for pixel-based unsupervised classification (Berman et al., 2014), and 9) circadian activity and spontaneous locomotion in 96-well plates with access to food. Each of these assays produced multiple behavioral measures for each individual fly. For example, flies behaving in the phototactic "LED Y-maze" (assay 6) are performing phototaxis and exploratory locomotion but yield several different behavioral measures, including: the number of choices made by passing through the choice-point of the Y-maze (a measure of total activity), the fraction of turns that are to the right, the fraction of turns that are toward the lit LED, the number of pauses in which

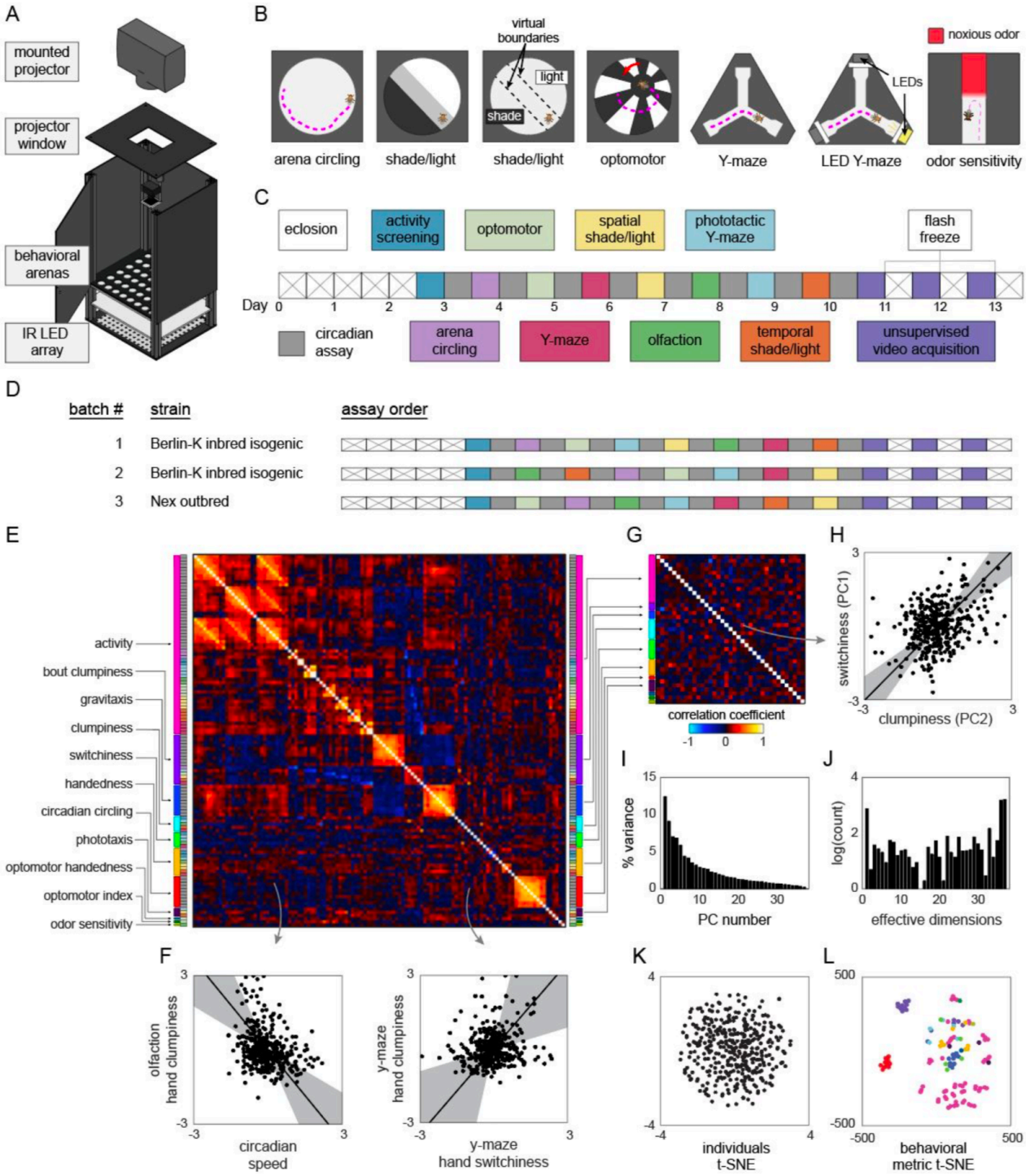


Figure 1 — Decathlon experimental design and structure of intragenotypic behavioral variation. A) Schematic of the imaging rig used for most Decathlon experiments. B) Schematics of the behavioral assays, illustrating the geometry of the arenas and stimulus structure. C) Timeline of the Decathlon experiment. Colors indicate the assays conducted on each day, half-black half-white blocks indicate the circadian assay and storage in 96-well plates. D) Timelines of the three Decathlon experiments, indicating the randomized order of assays 2-8. E) Full correlation matrix of all raw behavioral measures taken in the Decathlon. Colored blocks indicate blocks of measures we thought a priori might be correlated (outer blocks, text labels). Inner blocks indicate assay. F) Example scatter plots associated with measure correlations. Points are individual flies. Line is the best fit (PC1 of these points), grey region is the 95% confidence interval of the fit, as determined by bootstrap resampling. G) Distilled correlation matrix in which all correlated measures represent unexpected relationships. H) Example scatter plot from the distilled correlations. Plot elements as in F. I) Scree plot of the ranked, normalized eigenvalues, i.e., the % variance explained by each PC, of the distilled behavior matrix, versus PC #. J) Effective dimensionality spectrum (See text and Figure S9) for the distilled matrix. Height of bars indicates organization at that dimensionality. K) Points corresponding to individual flies non-linearly embedded using *t*-SNE from the 121-dimensional raw measure space to two dimensions. L) Points corresponding to behavioral measures non-linearly embedded using *t*-SNE from the 384-dimensional space of flies to two dimensions. Colors indicate groups of measures we expected a priori to be related.

the animal did not move, the average duration of pauses, etc. Thus, the total collection of behavioral measures across all assays per fly was quite large (up to 121), constituting a diverse, inclusive characterization of individual behavior. Each assay has a particular measure that captures the behavior it is primarily designed to assess (e.g., the fraction of turns toward the lit LED in the LED Y-maze). In control experiments, we confirmed that these primary measures are consistent across days within an individual (Figure S1) (i.e., they reflect persistent idiosyncrasies (Kain et al., 2012; Buchanan et al., 2015, Kain et al., 2015; Honegger et al., 2019)).

In order to obtain all of the behavior measures from each experimental animal, we combined our behavioral assays in a serial experimental pipeline lasting 13 consecutive days (Figure 1C), generally with one unique assay per day and continuous circadian imaging (assay 9) between assays. This pipeline begins with 3 day-old flies being loaded into the 96-well circadian imaging plates. Using a common behavioral platform, as much as possible, and storing flies between experiments in 96-well plates made maintaining the errorless identity of flies over the whole 13 day experiment substantially easier. Starting on day 3, daily assays began. On each day, flies were lightly anesthetized on an ice-chilled plate and aspirated, maintaining their identity, into the assay arrays. After the assay was completed (typically after 2hrs of recording) flies were again lightly anesthetized and returned to 96-well plates for renewed circadian imaging. On the first such day, flies were loaded into an array of circular arenas and imaged for total activity (in a version of assay 1). At this point, the most active 192 flies were retained for further testing. In preliminary experiments, we found that flies that were inactive at the beginning of the pipeline were very unlikely to produce substantial amounts of data over the rest of the pipeline. With the addition of this activity-screening assay, the total number of experiments was ten, and as each fly "competes" in all ten events, we refer to the entire pipeline as a Decathlon.

It is possible that the assay order has some effect on the recorded behavior measures. So we randomized the assay order between Decathlon implementations as much as possible (Figure 1D), subject to two restrictions: activity-screening was always the first assay, and high-resolution imaging for unsupervised analysis (assay 8) was always the last assay. (This assay has lower throughput, and three days were required to complete all 168 remaining flies. If this assay were performed earlier in the pipeline, it might introduce heterogeneity across subsequent assays.) When each fly completed its run through all Decathlon assays (i.e., over the three days of assay 8 imaging), it was flash-frozen in liquid nitrogen for RNA sequencing.

Behavioral Variability Among Inbred Flies has High Dimensionality and Sparse Pairwise Correlations.

To collect data that would reveal the structure of behavioral variation within a genotype, we conducted two Decathlons using highly inbred, nearly isogenic flies derived from the wild type strain Berlin-K (BSC#8522; Nöthel, 1981). We confirmed that this strain was, indeed, highly isogenic with genomic sequencing of individual animals, finding ~75 SNPs in the population across the entire genome (Figure S2). 115 flies completed the first Decathlon, and 176 the second. While we aimed to collect 121 measures per fly, a portion of values were missing, typically because flies did not meet assay-specific activity cutoffs. For subsequent analyses (see Figure S3 for a schematic of all analysis pipelines), it was sometimes necessary to have a complete data matrix. So we infilled missing values using the alternating least squares method, which, as judged by analyses of toy ground-truth data, performed better than mean-infilling (Figure S4). For the sake of maximal statistical power, we wanted to merge the data sets from the two Berlin-K^{iso} Decathlons. The correlation matrices of these two data sets were not identical, but were substantially more similar than expected by chance (Figure S5), implying that while there were inter-Decathlon effects, much of the same structure was present in each and merging them was justified. To do this, we z-score normalized the data points from each arena array/batch (within each Decathlon) across flies, thus eliminating any arena, assay, and Decathlon effects and enriching the data for contrasts between individuals. A grand data matrix was made by concatenating these batches (382 individuals x 121 behavior measures).

The full correlation matrix of this Berlin-K^{iso} data set is shown in Figure 1E. It contains a substantial amount of structure, indicating that large groups of behavioral measures covary. But the covariance of many pairs of measures in the matrix is not surprising. For example, almost all our assays generate some measure of locomotor activity (meanSpeed in circular arenas, number of turns in Y-mazes, meanSpeed in the olfactory tunnels, etc.), and one might expect that especially active flies in one assay will be especially active in another assay. Additional unsurprising structure in this matrix comes in identical measures recorded in each of the 11-13 circadian assays each fly completed. But, even in this first analysis, surprising correlations were evident. For example, flies with higher variation in the inter-turn interval in the olfactory assay ("clumpiness" in their olfactory turning) exhibited higher mean speed in the circadian assays, and flies with higher variation in inter-turn intervals in the Y-maze (clumpiness in their Y-maze turning) exhibited lower mutual information in the direction of subsequent turns in the Y-maze ("switchiness" in their Y-maze handedness) (Figure 1F). See below, Buchanan et al., 2015, and Akhund-Zade et al., 2019 for more about these measures.

To produce an exhaustive list of such non-trivial correlations, we distilled the grand correlation matrix to a smaller matrix (the "distilled matrix"; Figure 1G) in which two kinds of interesting relationships were revealed: 1) uncorrelated dimensions among measures for which we had a prior expectation of correlation (e.g., if meanSpeed in circular arenas is found to be uncorrelated with meanSpeed in olfactory tunnels), and 2) correlated dimensions among measures for which we had no prior expectation of correlation. Relationships of the former class were identified by enumerating, before we ran any correlation analyses, groups of measures we expected to be correlated ("a priori groups"; Figure 1E,G; Table S2). We looked for surprising independence within such groups by computing the principal components of data submatrices defined by the grouping (e.g., for the "activity" a priori group, by running PCA on the data set consisting of 382 individual flies, and 57 nominal measures of activity). We then replaced each a priori group submatrix with its projection onto its statistically significant PCs, as determined by a reshuffling analysis (see Methods, Figure S6). Some a priori groups largely

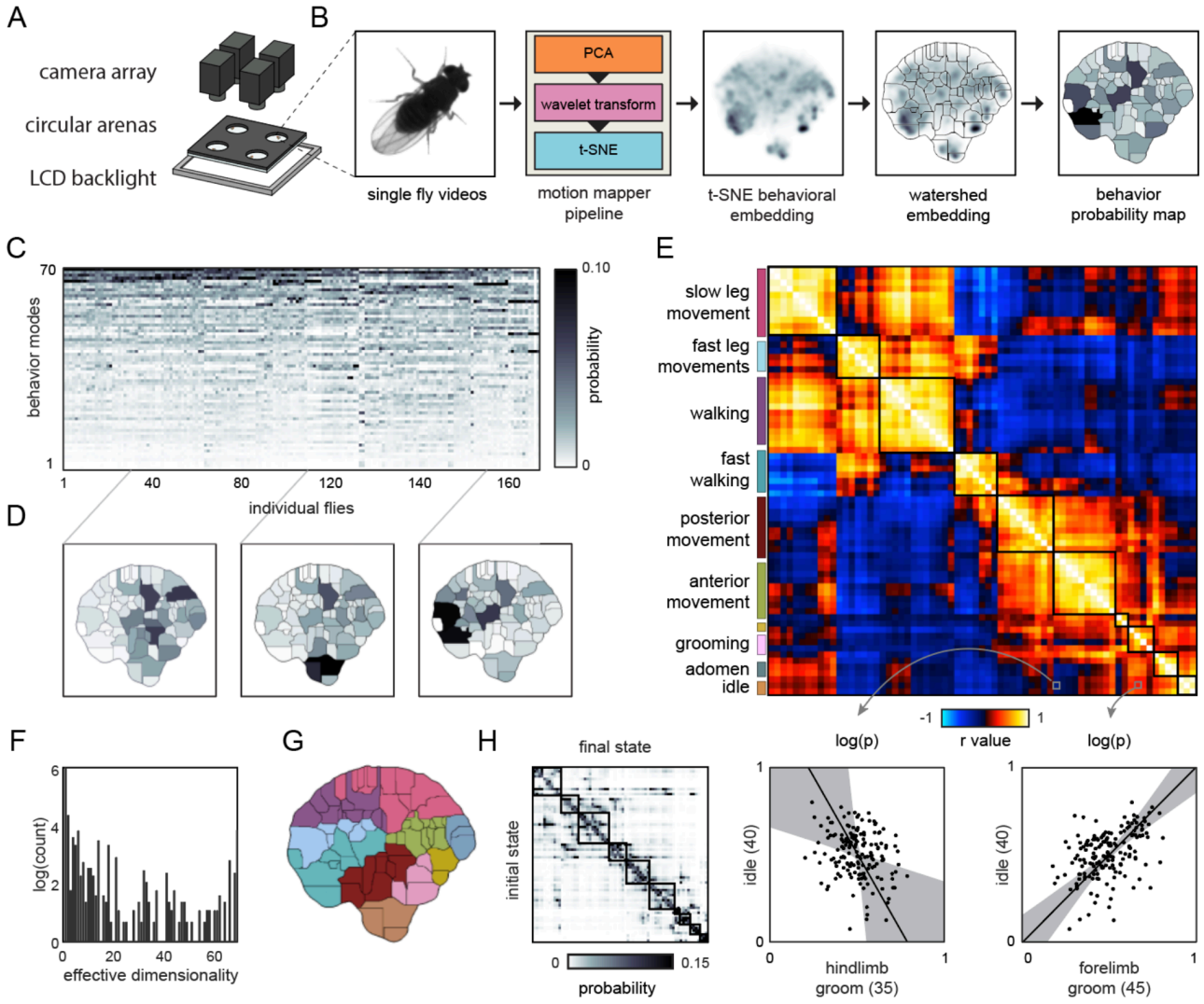


Figure 2 — Correlation structure of unsupervised behavioral classifications. A) Schematic of the four camera imaging rig used to acquire single fly videos. B) Overview of the data processing pipeline from single fly videos to behavioral probability maps. C) Behavioral classification probability density function (PDF) matrix. Columns correspond to behavioral PDFs for individual flies. D) Sample individual PDFs mapped to locations in *t*-SNE space. Discrete regions correspond to watersheds of the *t*-SNE embedded probability densities. E) Correlation matrix (top) for individual PDFs with rows and columns hierarchically clustered. Colored blocks indicate supervised labels applied to classifications post-hoc. Example scatter plots (bottom) of individual behavioral probabilities. Points correspond to probabilities for individual flies. Line is the best fit (PC1 of these points), grey region is the 95% confidence interval of the fit, as determined by bootstrap resampling. F) Connected components histogram of the thresholded PDF correlation matrix (see methods). G) Discrete behavioral map with individuals zones colored by supervised labels as in E. H) Transition probability matrix for behavioral classifications. Entries in the *i*th row and *j*th column correspond to the probability of transitioning from state *i* to state *j*. Blocks on the diagonal indicate clusters of supervised labels as in E.

matched our expectations, with relatively few independent dimensions among many measures (e.g., the gravitaxis group which had 10 measures and only 2 significant PCs), while others exhibited relatively many independent dimensions (e.g., the clumpiness group which had 5 measures and 5 significant PCs; see Figure S6 for all a priori group PCA analyses).

With a priori group submatrices represented in their respective significant PCs, the grand data matrix now contained 38 behavioral measures. Every significant correlation between behavioral measures at this point represents an unexpected element of structure of behavioral variation (Figure 1G). Given the high dimensionality of the full data set and the complex structure of the behavior measures in the distilled correlation matrix, we created an online data browser (*http://decathlon.debivort.org*) to explore the data and compare the alternative correlation matrices. The first impression of the distilled correlation matrix is that it is sparse. Most behavioral measures are uncorrelated or weakly correlated, meaning there are many independent dimensions of behavioral variation. However, 176 pairs of behaviors were significantly correlated at a false discovery rate of 38%, and the distribution of *p*-values for the entries in this matrix exhibits a clear enrichment of low values (Figure S7) indicating an enrichment of significant correlations. As an example, flies with high values in the first PC of the phototaxis a priori group tend to have high values in the second PC of the activity level a priori group. As another example, the first switchiness PC is positively correlated with the second clumpiness PC (Figure 1H; a relationship that is built, in part, on the positive correlation between the

y-maze hand clumpiness and y-maze hand switchiness, Figure 1F). Interpreting the loadings (Figure S6) of these PCs indicates that this is a correlation between olfactory tunnel turn direction switchiness and olfactory tunnel turn timing clumpiness. We detected a substantial number of correlations between different dimensions of switchiness and clumpiness (Figure S8), suggesting there are multiple couplings between these suites of traits.

Stepping back from specific pairwise correlations, we examined the overall geometry of behavioral variation. The full matrix contained 22 significant PCs, with PCs 1-3 explaining 9.3, 6.9 and 5.8% of the variance, respectively (the distilled matrix contained 16 significant PCs, with PCs 1-3 explaining 13.8, 10.0, and 7.7% of the variance, respectively; Figure 1I). But the amount of variance explained across PCs does not provide the full picture of how many independent dimensions of variation are present in a data set. For example, a correlation matrix can be organized at different scales/hierarchically, so there need not be a single number that characterizes effective dimensionality. We developed an "effective dimensionality spectral analysis" that assessed the continuous degree of organization of a data set across the continuous range of dimensionalities from 1 to *d*, the dimensionality of the data. Briefly, we thresholded the correlation matrix across a wide range of thresholds, and, treating these matrices as adjacency matrices, determined the number of connected components, recording how often *n* connected components were observed. See Methods and Figure S9. These spectra can be interpreted as follows: peaks at dimensionality = 1 indicate that all measures are coupled in a network of at least weak correlations; peaks at dimensionality = *d* indicate that all measures have some degree of independence; peaks in between these values indicate intermediate scales of organization. Multiple peaks are possible because these kinds of organization are not mutually exclusive. The effective dimension spectrum (Figure 1J) of the distilled Decathlon data set had peaks at 1 and *d* (37). There was also evidence for structure over the full range of intermediate dimensionalities. Overall, the organization is one of predominantly independent behaviors, with some sparse sets of behaviors correlated with continuously-varying strengths.

To assess how individual flies are distributed in behavior space, we embedded them from the 384 dimensional space into two dimensions using *t*-SNE (van der Maaten and Hinton, 2008). There appear to be no discrete clusters corresponding to "types" of flies. Instead, variation among flies appears continuously distributed around a single mode (Figure 1K). We also embedded behavioral measures as points from the 121 dimensional space of flies into two dimensions (Figure 1L). This confirmed that while our intuition for which sets of measures would be similar (the a priori groups) was right in many cases, measures we thought would be similar were often dissimilar across flies, and sometimes measures we did not anticipate being similar were (e.g., phototaxis and activity level).

Behaviors that Covary Among Individuals Tend to be Patterned Similarly in Time and Across the Body

With data from the second Decathlon, we characterized the structure of variation in a set of behaviors that was potentially exhaustive for one behavioral condition (free walking/motion in a 2d arena; Figure 2A). High-speed, high-resolution video was acquired for flies simultaneously in each of two rigs. Over three days, we acquired 13.5 GB of 200x200 px 100Hz videos centered on each fly as they behaved spontaneously over the course of 60 minutes. These frames were fed into an unsupervised analysis pipeline (Berman et al., 2014) that computed high-dimensional representations of these data in the time-frequency domain before embedding them in two dimensions and demarcating boundaries between 70 discrete modes of behavior (Figure 2B). The behavior of each fly was thus represented as one of 70 values at each frame. Flies exhibited a broadly similar probability distribution of performing each of these behaviors (Figure 2C), though there were conspicuous differences among individual patterns of behavior (Figure 2D).

The correlation matrix of behavioral modes identified in the unsupervised analysis was highly structured (Figure 2E; like the correlation matrix of the rest of the Decathlon measures, Figure 1E), appearing to have <10 major dimensions of variation, but also some evidence of organization in ~40 dimensions (Figure 2F). For this analysis, there was no equivalent of a priori groups of behavioral measures, as measures were not defined prior to the analysis. But, in examining sample movies of flies executing each of the 70 unsupervised behavioral modes (Movies S1-4), it was clear that highly correlated behavioral modes tended to reflect variations on the same type of behavior (e.g., walking) or behaviors performed on the same region of the body (e.g., anterior movements including eye and foreleg grooming; Figure 2G). In other words, individual flies that perform more eye grooming tend to perform more of other anterior behaviors. There were some correlations between behaviors implemented by disparate parts of the body. For example, flies that spent more time performing anterior grooming also spent more time performing slow leg movements (Figure 2G). The overall similarity of covarying behaviors was confirmed by defining groups of covarying behaviors and observing that they were associated with contiguous regions of the embedded behavioral map (Figure 2G). That is, behaviors whose prevalence covaries across individuals have similar time-frequency patterns across the body. Moreover, these clusters of covarying, contiguously embedded behaviors exhibited similar temporal transitions; behaviors that covary across individuals tend to precede specific sets of subsequent behaviors (Figure 2H). Thus, there appear to be couplings between the dimensions of behavioral variation across individuals, the domains of the body implementing behavior, and the temporal patterning of behaviors.

Thermogenetic Neural Perturbation Alters the Correlations Between Behaviors

To 1) confirm that the Decathlon experiments revealed biologically meaningful couplings between behaviors and 2) probe biological mechanisms potentially giving rise to behavioral correlations, we treated correlations in the Decathlon matrices as hypotheses to test using data from a thermogenetic neural circuit perturbation screen (Kakaria et al., 2019). Specifically, we focused on the many correlations between measures of turn timing clumpiness and turn direction switchiness (Figures 1H, S8). Before the Decathlon experiment, we had no reason to think these measures would be correlated, as one describes higher order structure in the timing of locomotor turns (clumpiness), and the other describes higher order structure in the direction of sequen-

tial turns (switchiness). Our Decathlon-derived prediction was that if perturbing a circuit element caused a change in clumpiness, it would tend to also cause a change in switchiness, in a consistent direction. We looked for such correlated changes when we inactivated or activated neurons in the Central Complex (Figure 3A-C), a cluster of neuropils involved in locomotor behaviors (Buchanan et al., 2015, Ofstad et al., 2011; Kottler et al., 2019) and heading estimation (Seelig and Jayaraman, 2015; Shah et al., 2017; Kakaria and de Bivort, 2017). We used a set of Gal4 lines (Wolff and Rubin, 2018), each of which targets a single cell type and that tile the entire Protocerebral Bridge (a Central Complex neuropil) (See Table 1), to express Shibire^{ts} (Kitamoto, 2001) or dTRPA1 (Hamada et al., 2008), thermogenetic reagents that block vesicular release and depolarize cells, respectively. As controls, we used flies heterozygous for the Gal4 lines, and lacking the effector transgenes.

Flies with these genotypes were loaded into Y-mazes for behavior imaging before, during and after a temperature ramp from the permissive temperature (23°C) to the restrictive temperature (32°C for dTRPA1 and 29°C for Shibire^{ts}) (Figure 3A). At the permissive temperature, we observed significant negative correlations in the average line clumpiness and switchiness of control and dTrpA1 expressing lines (Figure 3D,E). This suggests that the mechanisms which couple variation in switchiness and clumpiness within a genotype may also be at play across genotypes. Surprisingly, at the restrictive temperature, we saw significant positive correlations between clumpiness and switchiness in all three experimental treatments: control (Gal4/+), Gal4/Shibire^{ts} and Gal4/dTRPA1 lines. That this correlation appeared in controls suggests that temperature alone can selectively alter the function of circuit elements regulating both clumpiness and switchiness, effectively reversing their coupling. The dTrpA1- and Shibire^{ts}-expressing lines also showed this reversal, but to a lesser extent, suggesting that perturbing neurons in the central complex can block temperature-induced changes in the coupling of clumpiness and switchiness.

Individual Gene Expression Variation Correlates with Behavioral Variation

That thermogenetic manipulation can disrupt correlations between behavioral measures suggests that specific patterns of neural activity underlie the structure of behavioral variation. Such physiological variation could arise in stochastic variation in gene expression (Lin et al., 2016) in circuit elements. To test this hypothesis, we performed RNA sequencing on the heads of the flies at the end of the first Decathlon experiment (Figure 4A). We used Tm3'seq (Pallares et al., 2020) to make 3'-biased libraries for each individual animal. We quantified the expression of 17,470 genes in 101 individual flies. The expression profiles were strongly correlated across individuals (Figure 4B), but there was some evident variation across individuals (Figure 4C). To assess whether this variation was meaningful with respect to behavioral variation, we trained linear models (over 625,000 in total) to predict an individual's behavior measures from its transcriptional idiosyncrasies. Specifically, we fit simple linear models for each of our 97 behavior measures as a function of the 6,642 most highly expressed genes. The median model had an r^2 value of 0.008 and 5% of models predicted behavior at $r^2 >$ 0.135. Behavior measures varied greatly in their number of significant ($p<0.05$) gene predictors (Figure 4D), ranging from 147 to 1,172 genes.

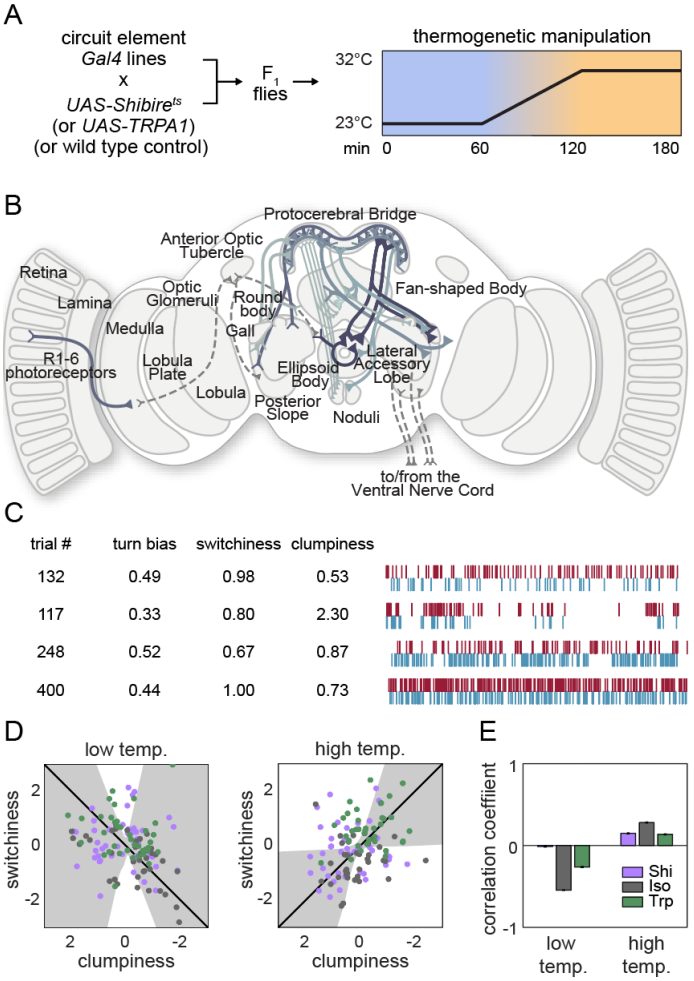


Figure 3 — Effect of thermogenetic neural perturbation on clumpiness and switchiness. A) Schematic of the experiment for testing the effects of neural perturbation on switchiness and clumpiness. Plot depicts the schema used to activate or silence neurons. Colored hash marks indicate periods when the light was on (cyan) or off (red). B) Diagram of the various neurons (solid lines) targeted by lines tested in the screen. See Table S3 for list. C) Example turn data from individuals exhibiting low and high clumpiness (upper rows) and low and high switchiness (bottom rows). D) Scatter plots of line average clumpiness and switchiness at the permissive (left) and restrictive (right) temperatures. Lines indicate the line of best fit and shaded regions indicate the 95% confidence of the fit as determined by bootstrap resampling. E) Average correlation coefficients for each effector type at the permissive and restrictive temperatures. Error bars indicate the 95% confidence interval as determined by bootstrap resampling.

After identifying genes that were predictive of each behavioral measure, we assessed whether they shared functional characteristics (Figure S11A) using KEGG pathway enrichment analysis (Mootha et al., 2003; Kanehisa and Goto, 2000; Yu et al., 2012). We independently performed KEGG enrichment analysis on each gene list to identify functional categories of genes overrepresented for any particular behavioral measure compared to the background list of 6,642 genes. Of this background, 1,982 genes (30%) were associated with one or more KEGG pathways. Across all behaviors, we found that 37 KEGG pathways were

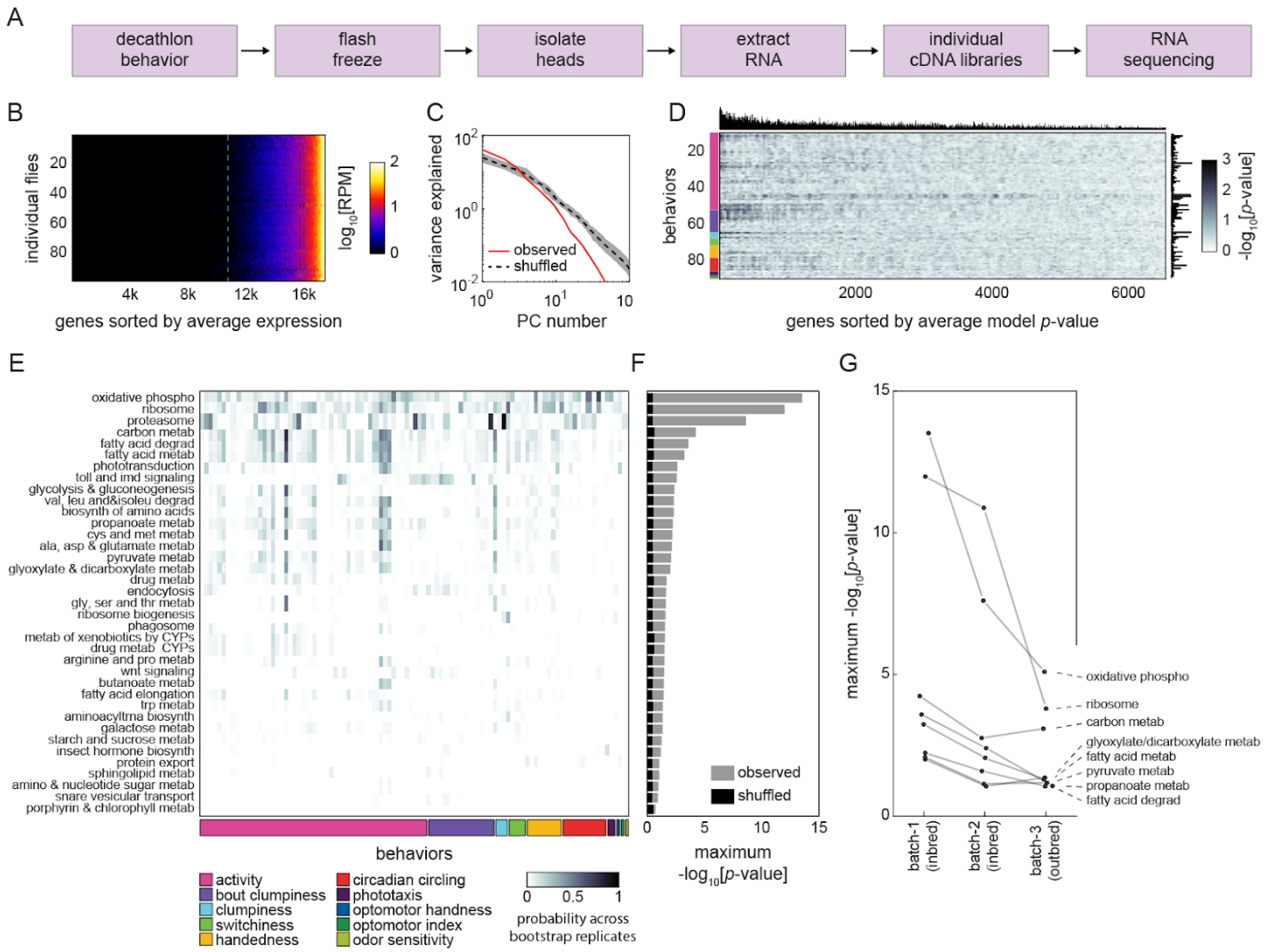


Figure 4 — Correlation between individual transcriptomes and behavioral biases. A) Steps for collecting transcriptomes from flies that have completed the Decathlon. B) Data matrix of individual head transcriptomes. Rows are individual flies (n = 98). Columns are 17,470 genes sorted by their mean expression across individuals. The dashed line indicates the mean expression cutoff at 10 reads per million (RPM), below which genes were excluded from analysis. C) Scree plots of the logged % variance explained for ranked principal components of gene expression variation, for observed (red) and shuffled (black) data. Shaded region corresponds to 95% CI as calculated by bootstrap resampling. D) Performance heatmap (-log p) of linear models predicting the behavior of individual flies from single gene expression. Colored bars (left) indicate a priori group identity of behavioral measures (rows). Bar graphs show the number of significant (p<0.05) models for each gene (top) and behavior (right). E) Heatmap showing the probability across bootstrap replicates of a KEGG pathway being significantly enriched in the list of predictive genes for a given behavior. F) Bar plot showing the average across bootstrap replicates of the maximum (across behaviors) negative log adjusted p-value of all enriched KEGG pathways. Color indicates results from observed (gray) or shuffled control (black) data. G) Average maximum adjusted -log p-value for enriched KEGG pathways common to all Decathlon iterations. Pathway labels (right) are ordered by batch 3 (outbred) -log adjusted p-value.

significantly enriched in the genes that predicted individual behavior measures (Figure 4E, S11G, Table S4). Many of these pathways were significantly enriched compared to shuffled controls and were enriched in multiple behavior measures (Figure 4F). We included features in the online Decathlon Data Browser to explore gene-behavior correlations (searching either by gene or by behavior) as well as KEGG pathway-behavior correlations. We repeated this experiment and analysis on flies of the second Decathlon experiment and found that the significance of pathway enrichments was highly correlated (r = 0.87) and substantially overlapping (55%) between Decathlon experiments (Figure 4G).

Genes related to cellular respiration, protein translation, and phototransduction were significantly enriched for multiple behavioral measures, a result that was highly robust to bootstrap resampling. Cellular respiration was the most common enrichment, significant in 11 of 97 behavioral measures, suggesting that variation in metabolic rate may be predictive of variation in many behaviors. Indeed, metabolic function was a common link between significant categories, with a total of 24 of 37 being related to metabolism. We also found 11 pathways (including two highly enriched categories: Ribosome and Proteasome) associated with protein turnover. While most enriched pathways were related to basic cellular processes, others (Hedgehog signaling, Wnt signaling, insect hormone biosynthesis, SNARE vesicular transport, and phototransduction) suggested roles for

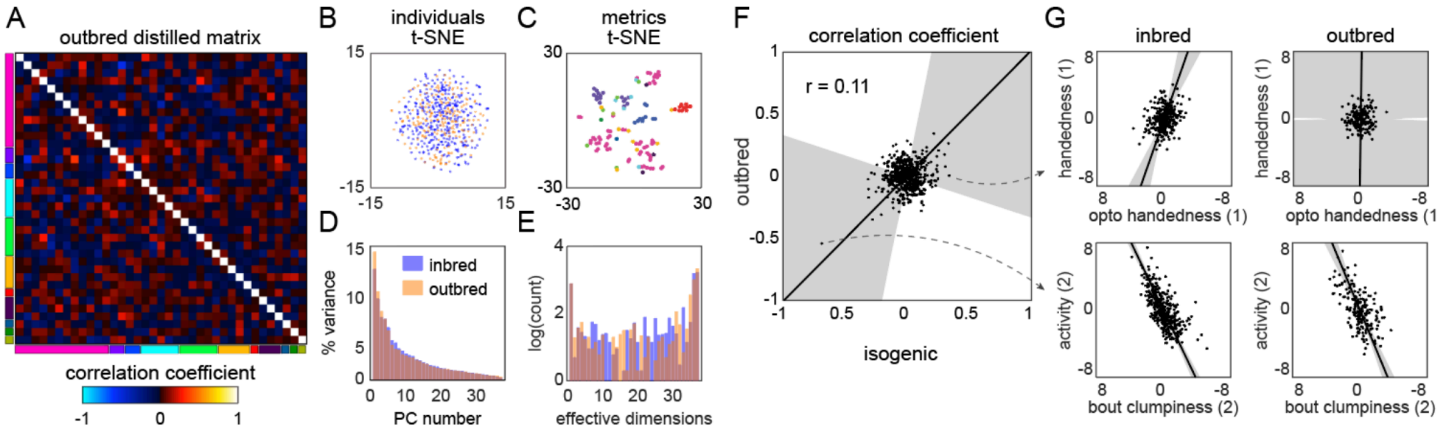


Figure 5 — Structure of behavioral variation in outbred flies. A) Distilled correlation matrix for outbred NEX flies. Colored blocks indicate a priori groups as described in figure 1. B) Points corresponding to individual flies non-linearly embedded using *t*-SNE from the 121-dimensional raw measure space to two dimensions. Color indicates whether the genotype is inbred (blue) or outbred (orange). C) Points corresponding to behavioral measures non-linearly embedded using *t*-SNE from the 192-dimensional space of flies to two dimensions. Colors correspond to a priori group. D) Scree plot of the ranked, normalized eigenvalues, i.e., the % variance explained by each PC, of the distilled covariance matrix, versus PC #. E) Log histogram of the effective dimensionality of the distilled matrix, as calculated by the number of connected components in the thresholded graph covariance matrix over 5,000 linearly spaced covariance thresholds (see methods). F) Scatter plot of the distilled matrix correlation coefficients for inbred and outbred flies. Points correspond to distilled matrix measure pairs. G) Example scatter plots of distilled matrix measure pairs for inbred (left) and outbred (right) flies. The rows of plots highlight a pair of measures with qualitatively different (top) and similar (bottom) correlations in inbred and outbred flies.

development and neuronal function in individual behavioral variation.

Behavioral Variability Has a Similar Structure in Inbred and Outbred Lines

If transcriptomic differences predict individual behavioral differences within a genotype, then the structure of behavioral variability might be very different in outbred populations, where transcriptomic differences are (presumably) much more substantial. We tested this by conducting a Decathlon experiment on outbred flies (n=192) from a synthetic genetic mapping population (Long et al., 2014). These animals were from a high (~100)-generation intercross population ("NEX"; seeded in the first generation by eight kinds of F1 heterozygotes produced by round-robin cross from eight inbred wild strains). A distilled correlation matrix of behavior measures (Figure 5A) was produced by the same method as above. At first glance, it appears qualitatively similar to the distilled correlation matrix from inbred animals (Figure 1G). This impression was confirmed in more formal comparisons of the structure of behavior in inbred and outbred populations. In inbred and outbred populations: 1) individuals do not fall into discrete clusters, as determined by *t*-SNE embedding of individuals as points (Figure 5B). Moreover inbred and outbred flies appear to lie on the same manifold in behavior measure space; 2) behavioral measures cluster according to their membership in a priori groups similarly in outbred (Figure 5C) and inbred (Figure 1L) populations; 3) the distribution of percent variance explained by principal component was similar (Figure 5D); and 4) there is a similar spectrum of covariance structure, with most measures independent, and a sparse network of correlations of varying strength (Figure 5E). We conducted transcriptomic sequencing and analysis on the heads of these outbred flies and found that, similar to the inbred flies, their behaviors were broadly correlated to gene expression (Figure S11C3). KEGG pathways enriched among these correlated genes largely overlapped (13 out of 22 total pathways) those of the inbred flies.

After determining that the overall structure of behavioral variation in inbred and outbred populations is similar, we asked whether it was also similar in specific correlations. There appears to be some similarity at this level (Figure 5F); the correlation coefficient between the inbred and outbred populations in the pairwise correlations between behavior measures is statistically significant (p=0.001), but low in magnitude (r=0.12). Examining specific pairs of scatter plots, it is clear that the correlations between specific behaviors are sometimes the same between the inbred and outbred animals, but sometimes not (Figure 5G). A caveat in interpreting apparent differences between the inbred and outbred matrices is that two qualities are different between the animals used in the respective experiments: the degree of genetic diversity, but also (necessarily) the genetic background of the flies.

Behavioral Variability has High Dimensionality Regardless of the Mechanistic Origins of Variation

Lastly, we examined how the correlation structure of behavior compared between sets of flies with variation coming from different sources. Specifically, we looked at four data sets: 1) the BABAM data set (Robie et al., 2017), in which measures were acquired from groups of flies behaving in open arenas, and variation came from the thermogenetic activation of 2,381 different sets of neurons (the first generation FlyLight Gal4 lines Jenette et al., 2012); 2) a *Drosophila* Genome Reference Panel (DGRP; Mackay et al., 2012) behavioral data set, in which measures were acquired in behavioral assays similar to the Decathlon experiments (sometimes manually, sometimes automatically), and variation came from the natural genetic variation between lines in the DGRP collection; 3) a DGRP physiological data set, in which measures are physiological or metabolic (e.g., body

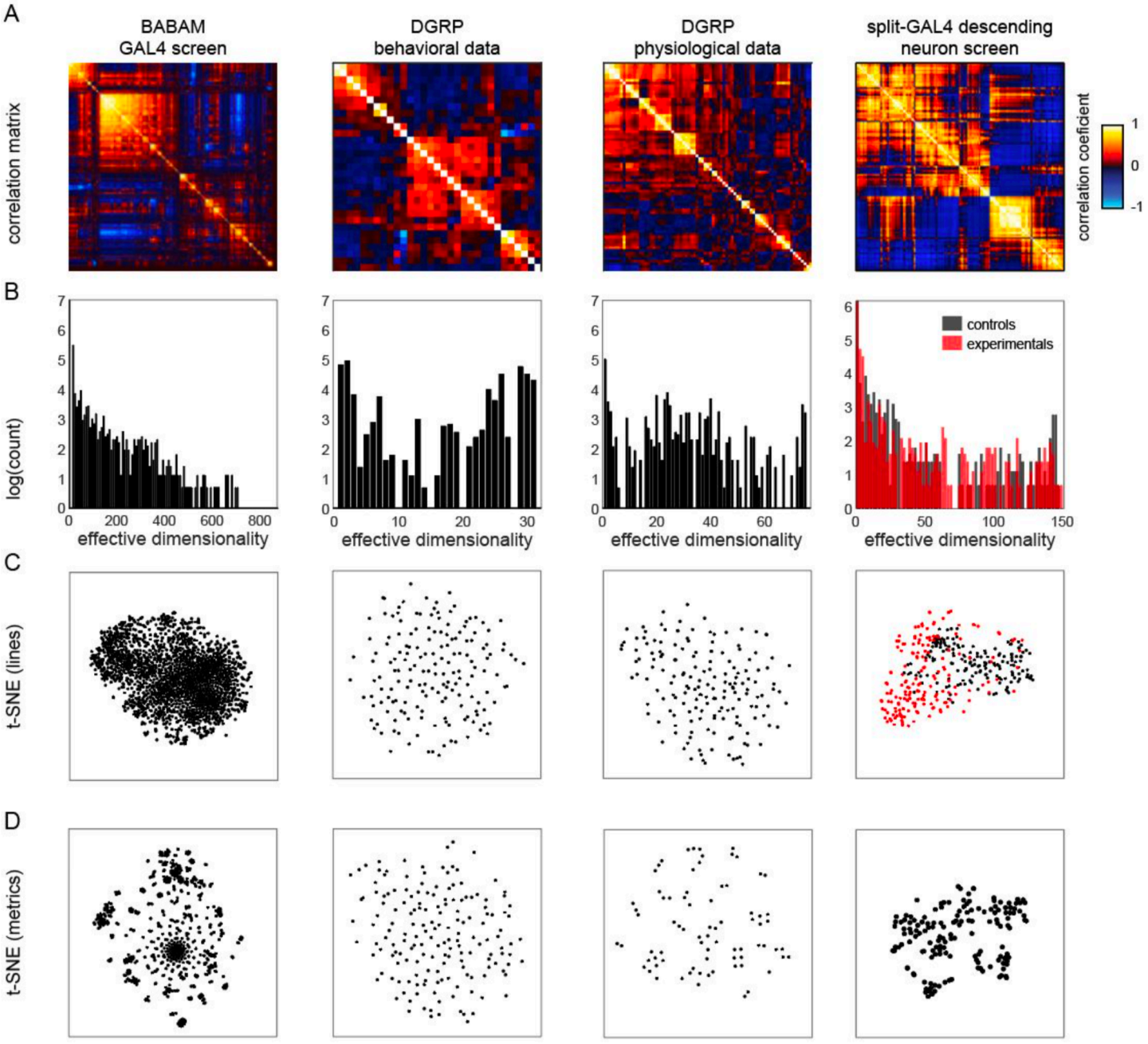


Figure 6 — Analysis of *Drosophila* behavioral covariation in other non-isogenic populations. A) Correlation matrices of previously published data sets. Rows correspond to analyses performed on each data set. From left to right, the data sets (columns) are as follows: line averages of supervised behavioral classifications following thermogenetic inactivation in the fly olympiad screen (Robie et al., 2017), line averages of behavioral phenotypic data from wild-type inbred lines in the *Drosophila* Genomic Reference Panel (DGRP) database, line averages of physiological phenotypic data from the DGRP database, line averages of the fold change in unsupervised behavioral classifications following optogenetic activation of descending neurons (Cande et al., 2018). B) Log histograms of the effective dimensionality of all data sets, as calculated by the number of connected components in the thresholded graph covariance matrix over 1,000 linearly spaced covariance thresholds (see methods). Color in the rightmost plots (B-D) indicates either control (Gal4 driver only) or experimental animals (Gal4 x dTrapA1). C) Points corresponding to lines non-linearly embedded using *t*-SNE from the D-dimensional raw measure space to two dimensions (from left to right, $d = 871, 31, 77, 151$). D) Points corresponding to lines non-linearly embedded using *t*-SNE from the *n*-dimensional raw measure space to two dimensions (from left to right, $n = 2083, 169, 169, 176$).

weight and glucose levels) and variation came from the natural genetic variation between lines in the DGRP collection; and 4) the split-Gal4 Descending Neuron (DN) data set (Cande et al., 2018) in which measures came from the same unsupervised cluster approach as Figure 2, and variation comes from the optogenetic activation of specific sets of descending neurons projecting from the brain to the ventral nerve cord (Namiki et al, 2018). We analyzed these data sets with the same tools we used to characterize the structure of behavioral variation in the Decathlon experiments.

All of these data sets show substantial structure in their correlation matrices (Figures 6A, S10). The BABAM and especially the DN correlation matrices contain numerous high correlation values, indicative of strong couplings between behaviors under these neuronal manipulations. The DGRP correlation matrices, especially the DGRP behavioral matrix, look more qualitatively similar to the Decathlon matrix, with lower, sparser correlations. This suggests that behavioral variation has coarsely similar structure whether variation arises intragenotypically (e.g., through stochastic variation in transcription; Figure 4B, Figure 1J), intergenotypically among outbred individuals (Figure 5E), or intergenotypically among inbred lines derived from wild populations (Figure 6A2). A caveat of this conclusion is that sparse correlation matrices can arise either from true, biological independence of behavior measures or from measurement error.

The effective dimensionality spectra of these matrices (Figure 6B) are similar in offering evidence of organization over a wide range of dimensionalities, including high dimensionalities. Only the BABAM spectrum has no power at the dimensionality of its raw count of measurements. The BABAM and DN spectra have a single predominant peak (at dimensionality = 1), suggesting that most measures belong to a single network of at least weak couplings. This is especially true in the BABAM data, and more true in the optogenetic experimental animals than controls in the DN data. The DGRP physiology data exhibits weak peaks at dimensionality = 1 and *d*, but also a broad peak between 20 and 40, suggesting that intergenotypic variation in physiology may have an intrinsic dimensionality in that range. The spectrum of the DGRP behavior data looks similar to that of the Decathlons, with peaks at dimensionality = 1 and *d*, and evidence for structure over the full range of intermediate dimensionalities. The distribution of individual lines in the space of DGRP behavior (and physiology) measures appears to be distributed single mode (Figure 6C2, C3), like individual flies in the Decathlon (Figure 1K). In contrast, there is some organization of individual lines in the BABAM and DN data sets, likely reflecting neuronal perturbations affecting multiple circuit elements mediating the same behavior(s) (e.g., multiple lines targeting the same neuropil). Measures fall into clusters in all of these data sets except the DGRP behavior measures (Figure 6D), which appear distributed around a single mode, perhaps reflecting the high dimensionality of behavior itself.

Discussion

Individuals exhibit different behaviors, even when they have the same genotype and have been reared in the same environment. These differences might covary, or lie on a manifold of specific geometry in behavior variation-space, but the structure of intragenotypic behavioral variation is uncharacterized. We designed a pipeline of ten behavioral assays (Figure 1), which collectively yielded up to 121 behavioral measures per individual animal. We also used unsupervised clustering to identify an additional 70 measures per individual based on a time-frequency analysis of high resolution video of the flies behaving spontaneously (Figure 2). These measures were the fly-specific rates of exhibiting each of the 70 unsupervised behavioral modes. All in all, across three 15 day Decathlon experiments, we collected 191 behavior measures from 576 flies. This allowed us to produce a full correlation matrix for all of the behavioral measures for the variation present in inbred animals grown in the lab (Figure 1E).

There is a well-developed theoretical framework for understanding the multivariate correlation structure of phenotypes. In quantitative genetics, G-matrices characterize the variance and covariance structure of phenotypes (be they behavioral, physiological, morphological) stemming from genetic differences among individuals or strains (Lynch and Walsh, 1998, Bruijning et al., 2020). These representations allow the quantitative prediction of responses to selection and constrain the combinations of phenotypes individuals can exhibit. As such, these representations are a key part of predicting the future trajectories of evolution. Just as the phenotypic variance can be parsed into genetic variance, environmental variance, GxE interaction variance etc., covariance can be similarly dissected (Charmantier et al., 2014, Berdal and Dochtermann, 2019). For example the classic model of phenotypic variance $V_P=V_G+V_E$ has a direct phenotypic covariance analogue: $Cov_P = Cov_G + Cov_E$. The last term (Cov_E) is further broken down into Temporary Environmental covariances and Permanent Environmental covariances (Cov_{PE}) that endure for the duration of observations (exactly like our measurements of individual behavioral bias). In flies, we have the potential to directly measure Cov_{PE} by rearing inbred animals in standardized lab environments, profiling their individual biases over a wide range of behavioral measures, and directly quantifying the variance and covariance of behavioral bias. This is significant for quantitative geneticists because a meta-analysis of behavioral traits indicates that across behaviors, ~23% of variance can be attributed to heritable factors (Dochtermann et al., 2019). This means that environmental factors, which include both deterministic effects and stochastic intragenotypic effects, explain ~77% of behavioral variance. Thus, characterizing the structure of Cov_{PE} will contribute to closing a significant gap in our understanding of the basis of behavioral diversity.

For ethologists and behavioral neuroscientists, this work yields a view of the geometry of intragenotypic behavioral variation, which can be thought of as an emergent product of developmental biological processes and the dynamic interaction of neural activity and animals' environment. From the full behavioral matrix, we made a so-called "distilled" matrix in which any significant correlation indicates a surprising new relation between behaviors (Figure 1G and S8). This form of the data minimizes duplicated measures of the same behavior, allowing us to cleanly analyze the geometry of behavioral variation. We found that behavioral measures were largely independent of each other, so that the main effective dimensionality of the behavioral space matched the number of behaviors measured. But a single number cannot fully characterize the dimensional organization of a correlation matrix; so we developed a spectral approach that examined the degree of organization across all possible dimensionalities in the data (Figure 1J, 3E, 6B, S10). This revealed a degree of organization at intermediate dimensionality corresponding to sparse correlations between specific pairs of behaviors (Figure 1F,H, S9). We found no evidence of discrete types of flies. Embedding data points corresponding to individual flies from the high dimensional space of individual biases into two dimensions produced a broad distribution around a single mode (Figure 1K).

One of the specific, surprising correlations we discovered was between "clumpiness" and "switchiness" (Figures 1F,H, 3, S8). These are slightly abstract, higher-order behavior measures, corresponding respectively to the burstiness of turn/action/decision timing and the degree of independence between consecutive binary choices. We had no a priori reason to expect these measures would be correlated, since one pertains to the structure of actions in time, and one pertains to the persistence of trial-to-trial biases. However, their linkage may reflect a shared role in controlling the higher order statistics of exploration. Sequences of behavior with clumps of bouts (either in time or in space) might contribute to fat-tailed distributions of dispersal that are advantageous over Brownian motion for foragers in environments of sparse resources (Bartumeus et al., 2002). Variation in switchiness and clumpiness across individuals might therefore reflect variation in multi-behavioral navigational strategies, perhaps as part of bet-hedging evolutionary strategy (Hopper, 1999; Kain et al., 2015). From a perspective of biological mechanism, the correlation between these two behaviors (or other pairs we discovered) could be established during development. Individual wiring (Mellert et al., 2016; Linneweber et al., 2020) or physiological variations in neurons that mediate more than one behavior could impart coupled changes to all such behaviors.

If such an explanation accounts for the correlation of clumpiness and switchiness, there may be shared neural circuit elements in the circuits controlling decision-timing and decision-bias. We tested this idea in a thermogenetic screen of circuit elements in the central complex, a brain region where heading-direction is represented (Seelig and Jayaraman, 2015) in a ring-attractor circuit (Kakaria and de Bivort, 2017). We found that increasing the temperature changed the sign of the correlation between switchiness and clumpiness (Figure 3D,E), suggesting that changes to brain physiology can alter correlation structure, though these effects might instead be caused by any of the many changes in neural state that accompany a temperature change. Interestingly, we also found that the effector-inducing temperature manipulation alone concomitantly changed clumpiness and switchiness in some lines. This suggests that potentially subtle alterations of circuit physiology (e.g., temperature shifts (Haddan and Marder, 2018) well within physiological limits) can affect the function of circuit elements governing multiple behaviors. In contrast, we found that thermogenetic manipulation of neurons in the central complex can diminish the effect of temperature shift on the correlation between clumpiness and switchiness. At the same time, the correlation of these behaviors went up significantly for the lines expressing Shibire^{ts} at the restrictive temperature. Thus extreme neural perturbations (such as those induced by thermogenetic manipulation) can apparently couple or decouple behaviors.

We included behavioral assays in the Decathlon pipeline under a number of constraints. The assays had to be high throughput, both in the number of flies that could be assayed and in measures being automatically acquired. Flies had to survive at high rates, and the measures had to be stable over multiple days (Figure S1), because the whole experiment lasted 15 days. Because not all behavior measures showed robust stability for this duration (all showed at least some day-to-day stability), the distilled Decathlon matrices likely represent an underestimate of the behavioral couplings exhibited over short periods (and perhaps an overestimate of life-long couplings, as flies can live more than a month). In the end, we employed a number of spontaneous locomotion assays and simple stimulus-evoked assays like odor-avoidance and phototaxis.

Light responses were measured in a number of assays (Table S1), specifically: the LED Y-maze (Werkhoven et al., 2019) (in which flies turned toward or away from lit LEDs in a rapid trial-by-trial format), the Spatial Shade-Light assay (in which flies chose to stand in lit or shaded regions of an arena that only changed every four minutes), and Temporal Shade-Light (in which the same luminance levels were used as the previous assay, but a fly experienced them by traveling into virtual zones which triggered the illumination of the whole arena at a particular luminance) (Figure 1B). These assays were potentially redundant, and we included this cluster of phototaxis assays in part as a positive control. However, the three phototactic measures we thought would be correlated a priori were, in fact, independent of each other, each being represented in the distilled correlation matrix (Figure 1G). This may reflect flies using different behavioral algorithms (Krakauer et al., 2017), implemented by non-overlapping circuits, to implement these behaviors. Indeed, independence between behavior measures was the typical observation. This also suggests that we have not come close to sampling the full dimensionality of intragenotypic behavioral variation; if we were able to add more measurements to the experiment, they too would likely be independent.

To address potential biases in our sampling of assay and measure space, we performed an unsupervised analysis (Berman et al., 2014, Cande et al., 2018) of flies walking spontaneously in arenas. This approach has the potential to identify all the modes of behavior exhibited in that context. Moreover, because the unsupervised algorithm is fundamentally a clustering algorithm, it does not necessarily return a definitive number of clusters/behavioral modes (with more data, it can find increasingly more clusters). Because we can also extract second-order behavioral measures from this approach, such as Markov-transition rates between modes, this approach has the potential to yield a huge number of measures. In the end we were conservative in the number of measurements we chose to work with, matching it to the same order of magnitude as the number of flies we tested. The correlation matrix for the unsupervised behavior measures featured stronger correlations than the distilled Decathlon matrices (Figure 2E). Yet it had a similar effective dimensionality spectrum, indicating many independent dimensions of variation and that not all behavior modes had been sampled (Figure 2F).

Interestingly, the blocks of structure in the correlation matrix aligned, to some extent, with the blocks of structure in the Markov transition matrix of these behavioral measures. This suggests that behaviors mediated by non-overlapping circuits (those that vary independently across individuals) more rarely transition to each other over time. Conversely, behaviors mediated by overlapping circuitry are likely to follow each other sequentially. This may reflect the influence of internal states (Calhoun et al., 2019), with an internal state jointly determining what subset of overlapping neurons drives behaviors that are appropriate to string together in succession (e.g., Seeds et al., 2014). We did not assess the day-to-day persistence of behavioral modes identified in the unsupervised analysis, so the observed

variation across flies could reflect moods rather than permanent personalities. However, previous supervised (Kain et al., 2013) and unsupervised (Todd et al., 2017) analyses of spontaneous microbehaviors similar to those identified by unsupervised approaches have found such behaviors to persist across days.

We investigated whether individual variation in transcript abundance would predict individual behavioral variation (Figure 4). At the end of the Decathlon, flies were flash frozen and their heads were RNA sequenced. We fit linear models for thousands of gene-behavior measure pairs to identify a set of genes that predict behavioral variation. Some measures had many gene predictors, while others had none. Gene function enrichment analysis revealed that variation in the expression of genes involved in respiration and other kinds of metabolism predicted variation in many behavioral measures. Genes involved in neuronal and developmental processes were also enriched among predictors of individual behavior. These two functional categories are likely linked as there are strong causal couplings between metabolism and neural activity (Mann et al., 2020). Variation in behavioral measures without expression correlates may not have mechanistic origins in transcriptional variation, or the genes that do determine these behaviors may be not expressed in adults, expressed at low levels, or expressed in too few cells to detect in bulk head tissue.

Increasing transcriptional variation by adding genetic variation had the potential to change the correlation structure of behavioral measures. To our surprise, the distilled correlation matrix of outbred flies was qualitatively similar to that of our original inbred Decathlon (Figure 5). Both outbred and inbred matrices were dominated by independent axes of variation and sparse correlations between axes, with rough agreement in the specific pairwise correlations between these two data sets. These two data sets may have differed in their absolute variances (while appearing qualitatively similar in their covariances), but the normalization steps we took to resolve inter-Decathlon and assay batch effects precluded easy assessment of this possibility. That outbred and inbred variation had qualitatively similar structures raises the possibility that the same kinds of biological fluctuations underlie behavioral variation in populations of each of these kinds.

We finally examined the structure of behavioral variation in collections of flies where variation came from three additional sources: thermogenetic activation of 2,205 sets of neurons across the brain (Robie et al., 2017), optogenetic activation of 176 sparse populations of descending neurons connecting the brain to the motor centers of the ventral nerve cord (Cande et al., 2018), and variation in genotype across ~200 inbred strains derived from wild flies (Mackay et al., 2012). Overall, all of these datasets exhibited high degrees of independence in their behavioral measures. The structure of behavioral variation in the neural activation data sets was somewhat different from that of inbred and outbred flies, with a dimensionality smaller than that of the number of measures, and substantially more organization in lower effective dimensionalities (Figure 6A, B). These data sets also showed clustering of individual flies in behavior space (Figure 6C). Behavioral variation across the inbred strains derived from wild flies was organized qualitatively similarly to the variation across individual flies in inbred and outbred populations, again suggesting that the biological fluctuations across genotypes mirror those within a genotype as respects the coordination of behavior.

This work represents the most complete characterization to date of the structure of behavioral variation within a genotype. We found that there are not discrete types of flies, and there are many independent dimensions of behavioral variation. Moreover, the similar organization of biological variation within and among genotypes suggests that fluctuations in the same biological processes underpin behavioral variation at both of these levels. Elucidating the causal molecular and genetic fluctuations underpinning intragenotypic variation and covariation will be a challenging and revealing future research direction.

Methods

Data and Code Availability

Software

All software for analysis and figures in this paper as well as documentation are available at: *http://lab.debivort.org/structure-of-behavioral-variability/*. Static DOI for this code available at: *https://doi.org/10.5281/zenodo.4110049*. Actively maintained versions of the analysis software available at: *https://github.com/de-Bivort-Lab/decathlon*.

MARGO, animal tracking and stimulus delivery software: *https://github.com/de-Bivort-Lab/margo*.

Decathlon Data Browser, interactive web application for exploring data: *http://decathlon.debivort.org*.

Data

Data such as: behavioral measures, unsupervised embedding PDFs, RNAseq read counts, and formatted versions of the BABAM, DGRP phenotypic, and DN screen data are available at *http://lab.debivort.org/structure-of-behavioral-variability/*. Static DOI for these data in addition to the unsupervised classification embedding time series data is available at: *ttps://doi.org/10.5281/zenodo.4110049*. Raw assay tracking and time series data and raw video data for unsupervised behavioral classification are available upon request.

Fly strains

All Decathlon experiments were performed on virgin female fruit flies derived from inbred wild type Berlin-K (BSC# 8522) isogenic flies or a custom outbred flies (NEX) (REFEF). Prior to Decathlon testing, a lineage of Berlin-K isogenic flies (Berlin-K^{iso}) were selected for robustness from pool of inbred lineages as the result of a short screen designed to mimic the stresses of the Decathlon assay battery. All inbred lineages tested were derived from three wild type strains: Berlin-K (BK), Canton-S (CS), and Oregon-R (OR). For each wild type strain, six virgin females were dispensed into separate vials and paired with a single male fly to establish separate lineages. Parental flies were removed from their vials after 2 days to separate them their progeny. For

all successive generations after the first generation, the six virgin females were picked from the three vials (2 from each) with the highest number of progeny to select for fecundity and overall ease of maintenance. Virgins were then dispensed into separate vials with full sibling males for inbreeding. After 13 generations of inbreeding this way the resulting lineages (6 total for each strain) were screened for overall activity level via behavioral testing on the Y-maze assay in cohorts of 120 flies from each lineage. For each strain, the lineage with the highest number of choices (turns) in the Y-maze from was kept for further testing to improve overall sampling in behavioral experiments, resulting in a single lineage from each strain: Berlin-K^{iso}, Canton-S^{iso}, and Oregon-R^{iso}.

To screen robustness to the stresses of multiple days of behavioral testing such as periodic food deprivation and repeated cold anesthetization, cohorts of 96 flies from the resulting inbred lines were singly housed and repeatedly tested on the Y-maze assay over a 3 day period. Berlin-K^{iso} and Canton-S^{iso} were selected as the lines with lowest and second lowest (respectively) mortality after 3 days of testing. To introduce hybrid vigor and reduce the deleterious effects of inbreeding, four F1 hybrid inbred strains were generated by crossing each combination of parental sex and inbred strain (e.g., ♂ Berlin-K^{iso} x ♀ Canton-S^{iso}, ♀ Berlin-K^{iso} x ♂ Canton-S^{iso} etc) and screened for mortality in the 3 day experiment described above. Contrary to our expectation that increased heterozygosity in the F1 hybrid lines would result in more robust flies and lower mortality, F1 flies displayed intermediate mortality and activity level (choice number) between Berlin-K^{iso} and Canton-S^{iso} parentals. Thus Berlin-K^{iso} flies were ultimately selected for the Decathlon screen.

Fly handling

Parental flies and Decathlon flies (prior to individual housing) were raised on CalTech formula cornmeal media under 12 h/12 h light and dark cycle in an incubator at 25˚C and 70% humidity. A single cohort of virgin Decathlon flies were collected over an 8hr period (10AM-6PM) and stored in group housing up to a maximum of 288 flies. On the following day, flies were anesthetized using carbon dioxide (CO2) and were aspirated into custom multiwell plates for individual housing (FlySorter LLC, Seattle WA) on cornmeal media. Individually housed flies were stored in a custom imaging box for circadian activity measurements (circadian chamber) on a 12 h light and dark cycle (10AM-10PM) which was housed inside of an environmentally controlled room at 23˚C and 40% humidity. Food media was replaced and individual housing trays were cleaned every 2 days while flies were in behavioral testing to keep food moisturized and prevent microbial buildup.

Following the start of Decathlon (3 days post-eclosion), flies were removed from the circadian chamber each day for behavioral testing between 11AM and 2PM. Prior to all behavioral experiments (excluding olfaction) flies were cold anesthetized by transferring individual housing plates and food trays to an ice-chilled aluminum block in a 4˚C for 10 minutes. This time was minimum time required to reliably induce chill coma in most flies given the mass and low thermal conductivity of the cornmeal media separating the flies and the chill block. Once anesthetized, flies were individually transferred via aspiration to room-temperature custom behavioral arenas (without food) where they quickly awoke (typically 1-10 sec) after transfer. After all flies were transferred, they were given an additional 20 minutes to recover prior to behavioral testing. Flies were then tested in custom behavioral platforms, lasting as little as 15 min (olfaction only) and as much as 2 hrs (typical). Following testing, flies were returned to individual housing via aspiration either directly (olfaction only) or after re-anesthetizing by transferring the custom behavioral arenas to an ice-chilled aluminum block for 10 minutes. Once all flies were returned, the individual housing plates were returned to the circadian chamber between 2-5PM.

Behavioral Assay fabrication

Unless otherwise specified, all tracking was conducted in custom imaging boxes constructed with laser-cut acrylic and aluminum rails. Schematics of custom behavioral arenas and behavioral boxes were designed in AutoCAD 2014 (AutoDesk). Arena parts were laser-cut from black and clear acrylic and were formed as stacked layers joined with Plastruct plastic weld. Arena floors were made from sandblasted clear acrylic and arena lids were made from custom-cut clear eighth inch acrylic. Schematics for behavioral boxes and behavioral arenas can be found on the de Bivort Lab schematics repository (https://github.com/de-Bivort-Lab/dblab-schematics). Illumination was provided by dual-channel white and infrared LED array panels mounted at the base (Part BK3301, Knema LLC, Shreveport, LA). Adjacent pairs of white and infrared LEDs were arrayed in a 14×14 grid spaced 2.2cm apart. White and infrared LEDs were wired for independent control by MOSFET transistors and a Teensy 3.2 microcontroller. Two sand-blasted clear acrylic diffusers were placed in between the illuminator and the behavioral arena for smooth backlighting.

For circadian experiments, flies were housed and imaged in individual fly storage units (FlyPlates) from FlySorter, LLC. Circadian imaging boxes consisted of a small (6”x6”x18”) enclosed behavioral box with a dual-channel LED illuminator on a 12:12hr light and dark cycle programmatically controlled via MARGO behavioral tracking software (https://github.com/de-Bivort-Lab/margo). A heat-sink was affixed to the underside of the illuminator panel (outside of the box) and fan were used to ensure the interior of the boxes remained consistent with the temperature of the environmental room.

The olfactory sensitivity assay assay was performed in a custom behavioral chamber modeled on Claridge-Chang et al., 2009. The apparatus consisted of 15 parallel tunnels constructed from Accura 60 plastic using stereolithography (In’Tech Industries) fabrication. Stainless steel hypo tubing (Small Parts) was used to connect the apparatus with (ID: 0.7mm). Odorized or clean air was delivered via teflon odor tubes to inlet ports at each end of the tunnel and streams vented to the room through exhaust ports in the center forming a sharp choice zone. An active vacuum was not applied to the exhaust ports, and the tunnels operated close to atmospheric pressure throughout the experiment. A clear acrylic lid was clamped in place above the apparatus to ensure an air-tight seal during odor presentation. Air dilutions could be made independently for each side of the apparatus. A custom 15-way PEEK manifold was used on each side to split the odorized

flow equally between 15 tunnel inlets. A final valve (SH360T041; NResearch) was used immediately upstream of each manifold to quickly switch between pure dehumidified air and the odorized stream.

Phototactic stimuli were delivered with a custom 12"x12" PCB designed in Express PCB CAD software and manufactured by ExpressPCB. Briefly, the PCB was designed to power and independently control 216 white light LEDs (11.34mW±4.3μW at 20ma) geometrically arranged to match the maze arm ends of an array of Y-shaped behavioral arenas. LED intensity control was provided over USB via board interface with a teensy 3.2 microcontroller and constant current driver boards (Adafruit 24-Channel TLC5947 LED Driver). Holes in the footprint shape of each individual arena were custom cut into the PCB with water jet cutter to provide infrared backlighting.

Visual stimuli in the spatial shade-light, temporal shade-light, and optomotor assays were delivered in a behavior box modified to accommodate stimulus delivery with an overhead projector (Optoma S310e DLP). The DLP color wheel was removed reduce apparent flicker to the flies caused by low sensitivity to red light in flies. Stimuli were accurately targeted to individual flies by creating a registration mapping between the projector and the tracking camera using previously described method using MARGO tracking software. Briefly, a 2D polynomial registration model of a projector targeting coordinates of a small (10px) was fitted to coordinates obtained by tracking the dot as it was rastered over the projector display field. All visual stimuli were crafted and displayed with PsychToolbox.

Single fly videos used in the motionmapper unsupervised classification pipeline were acquired in behavioral boxes as described above. We replaced the dual channel LED illuminator with an LCD screen backlight to reduce apparent non-uniformities in the videos. Behavioral arenas consisted of custom thermoformed clear PTEG lid with sloping sides to prevent wall walking. Lids were coated with sigmacote (Sigma-Aldrich SL2-100ML) to prevent flies from walking on the ceiling and were clamped in place to a 0.25" tempered glass base via custom acrylic 2x2 array of arenas.

Behavioral Tracking

Flies were imaged with overhead tracking cameras at varying resolutions and frame rates. Unless otherwise specified, images for behavior tracking were acquired at 10Hz with a 1280 x 1024 pixel BlackFly GigE camera (PointGrey BFLY-U3-13S2M-CS) fitted with a Fujinon YV2.8×2.8SA-2, 2.8mm-8mm, 1/3", CS mount lens. Circadian and optomotor experiments were conducted with 3Hz and 60Hz imaging respectively. Tracking images for the odor sensitivity assay were acquired at 1328 x 1048 pixel at 23Hz (PointGrey FMVU-13S2C). Single fly videos for unsupervised classification were acquired.

Fly tracking and stimulus control for all experiments (excluding unsupervised single fly imaging) were programmed in MATLAB and implemented with MARGO (Werkhoven et al., 2019). The MARGO tracking algorithm has been previously described in detail. Briefly, binary foreground blobs were segmented from a thresholded difference image computed by subtracting each frame from an estimate of the background. A rolling model of the background was computed as the median image of the rolling stack of background sample images. Every 2min, a new sample image of the background was acquired and replaced the oldest image in the stack. Regions of interest (ROIs) were defined to encompass only a single behavioral arena (i.e., one fly), and the tracking algorithm proceeded independently for each ROI. The tracking threshold used to segment binary foreground blobs from the background was manually set prior to tracking and was independent for each experiment. Blobs below a minimum or maximum area (also defined manually for each experiment) were excluded from tracking in each frame. After filtering, the centroid trace of each ROI was updated with the position of the blob with shortest distance to the last known (non-NaN) position of each trace. The centroid trace of any ROI with no detected blobs in any given frame received no position (i.e., NaN). Output measurements of speed and area were converted from pixels/sec and pixels^2 to mm/sec and mm^2, respectively, by measuring the length a known landmark size (e.g., arena diameter) prior to tracking. Fisheye distortion of camera lenses were modeled and corrected via MATLAB's camera calibrator app.

For the acquisition of single fly videos used in the motionmapper unsupervised classification pipeline, flies were tracked in real-time to reduce video file size by extracting a 200x200 pixel region around the fly centroid. Tracking performed by computing the centroid of the entire difference image acquired calculated via the method described above. Tracking was coordinated simultaneously for four 1280x1024 pixel resolution cameras (PointGrey CM3-U3-13Y3M-CS) at 100Hz with a custom LabView script obtained from the J. Shaevitz lab (personal communication) which was lightly modified to add background subtraction to the tracking.

Decathlon Experimental Design

Flies underwent behavioral testing from 3 days post-eclosion up to a maximum of 14 days post-eclosion depending on when unsupervised classification was performed. For each Decathlon experiment, behavioral testing began with activity screen of 288 flies in the Y-maze assay. Flies were then sorted by number turns made during the activity screen. The top 192 flies with the highest activity level were selected for testing in the remainder of the Decathlon and were transferred to individual housing on day 3 post eclosion for overnight circadian behavior profiling. On subsequent days, all flies were anesthetized as described above and transferred from individual housing into a behavioral assay each day (from days 3-11 post-eclosion) and were returned individual housing for overnight circadian behavioral measurement. Flies were tested in cohorts dependent on the throughput capacity (described for each assay below) of the assay run on any given day. Although the range of testing times for all animals on some days was relatively broad (all experiments were conducted between 11AM-5PM), no behavioral experiment lasted more than 2hrs in duration, meaning that individual cohorts typically spent more than 3hrs (2hr experiment + 2x anesthetization and transferring) off of food media. The order of individual assays (shown in figure 1) was randomized at the start of each Decathlon with the exception of activity screening and unsupervised imaging which occurred at the beginning and end of each Decathlon (respectively) for logistical feasibility. Descriptions and imple-

mentation details of the Decathlon behavioral assays are provided below and are summarized in supplemental Table 1.

After day 11 post-eclosion, remaining living flies were split into three roughly equal-sized cohorts (approximately 56 flies each) for unsupervised behavioral imaging which were imaged in groups of 4 flies between 10AM-6PM on subsequent days. Cohorts of flies were flash-frozen on liquid nitrogen immediately following unsupervised behavioral imaging and prepped for RNA-seq as described above.

Arena Circling and Circadian Assays

Both the arena circling and circadian assays consisted of exploration of a circular arena. Although a near identical list of measures of locomotor handedness, speed, and movement bout dynamics were recorded in each assay (circadian includes gravitaxis also), the assays are distinguished by their background light level, duration, behavioral arena, and access to food. The arena circling assay was conducted over 2 h with constant light in a wide (28 mm diameter) and shallow (1.6 mm depth) arena without food. The circadian assay was conducted overnight (20-21 h) on a 12:12 h light and dark cycle (10AM:10PM) with each fly in a narrow (6.8 mm) and deep (10.6 mm) well of a 96 multiwell plate. The depth of the arena in addition to the difference in arena ceiling (sigmacoted acrylic in arena circling and uncoated plastic mesh circadian) add a vertical and body orientation dimension to the circadian assay not present in arena circling. This feature not only adds a measure of floor vs. ceiling preference (i.e., gravitaxis) to the circadian assay but also potentially confounds measures of circling directionality due to the fact that flies can circle the arena right-side up on the floor, upside-down on the ceiling, or sideways on the walls.

LED Y-Maze Assay

Individual flies explored symmetrical Y-shaped arenas with LEDs at the end of each arm. For all arenas in parallel, real-time tracking detected which arm the fly was in at each frame. At the start of each trial, an LED was randomly turned on in one of the unoccupied arms. Once flies walked into one of these two new arms, all the LEDs in the arena were turned off and a new trial was initiated by randomly turning on an LED in one of the newly unoccupied arms. This process is repeated for each fly independently over 2 hrs. Turns were scored as positively (toward a lit LED) or negatively (toward an unlit LED) phototactic. LEDs were either fully off (OFF) or were driven at maximum intensity (ON). No intermediate LED intensity values were tested. In addition to the phototactic bias, measures of locomotor handedness peed, choice timing, and choice sequence were also recorded.

Odor Sensitivity Assay

In the odor sensitivity assay, flies explored linear tunnels where half of the tunnel contained an odor (typically noxious) and the other contained no odor. Odor ports positioned at either end of the tunnels delivered either odor or odorless air. Vents positioned at the midpoint of each tunnel form a sharp boundary between the two air streams. Flies were presented with a 3min baseline period with no odor on either side followed by a 12min experimental period with half odor and half no odor. Sensitivity to the odor was estimated by measuring the fraction of occupancy time spent in the odor half of the tunnel. Flies showed a weak aversion to the designated odor side during baseline measurement, prior to the presentation of odor (0.47±0.30 mean pre-odor occupancy). Flies displayed strong initial aversion (0.17±0.22 mean first half odor occupancy) to the side of the tunnels with the odor, which slightly diminished (0.22±0.23 mean second half occupancy) over the stimulus duration. In addition to odor occupancy, locomotor handedness was scored as the average direction of heading direction reversals on the long axis of the tunnels. In addition to the above measures, speed, turn timing, and turn direction sequence were also recorded.

Optomotor Assay

In this assay, optomotor stimuli were centered on the bodies of flies by projecting them onto the floor of their arenas in which they are walking freely. These stimuli consisted of a maximally-contrasting (black = 0, white = 1, min. power = 79µW±61nW, max. power = 2.4mW±4.7µW) rotating pinwheel (spatial frequency = 0.18 cycles/deg, rotational speed = 320 deg/sec) and typically evoked a turn in the direction of the rotation to stabilize the visual motion. The pinwheel center followed the position of the fly as it moved so that the only apparent motion of the stimulus is rotational motion around the body. Therefor, the stimulus was closed-loop with respect to position and open-loop with respect to rotation velocity. We observed that optomotor responses could be reliably elicited, provided individuals were already moving when the pinwheel was initiated. We therefore only presented the pinwheel when: 1) flies were moving 2), a minimum inter-trial interval (2sec) had passed to prevent behavioral responses from adapting, and 3) flies exceeded a minimum distance from the edge of the arena to ensure that the stimulus occupied a significant portion of the animal's FoV. Over 2hrs of testing, an optomotor index was calculated for each fly as the fraction (normalized to [-1,1]) of body angle change that occurred in the same direction as the stimulus rotation over the duration of the stimulus. On average, flies displayed reliable optomotor responses (mean=0.49±0.16) when stimulated. In addition to optomotor index measures of fly activity, movement bout dynamics, and locomotor handedness were also recorded.

Spatial Light-Shade Assay

Ambient light preference is estimated in this assay by projecting a circular stimulus with one side fully bright (intensity = 1) and the other side fully dark (intensity=0), with the two regions separated by a small (10% of the arena diameter) boundary zone of intermediate brightness (intensity = 0.5) (min. power = 79µW±61nW, max. power = 2.4mW±4.7µW). A light occupancy measure was calculated for each fly as the fraction of time spent in the lit region divided by the total time spent in both the lit and unlit regions. Time spent in the intermediate boundary zone was scored as no preference and was excluded from the calculation. Each 2 hr experiment was divided into 15 stimulus cycles with each cycle consisting of an alternating 4min baseline block where each arena was unlit and a 4min experimental block where shade-light stimuli were targeted to each arena. At the start of each block, each stimulus was rotated to center the intermediate zones (or virtual intermediate zone in the baseline block) over the flies bodies. In effect, this required flies to move from intermediate zone during the block to express any prefer-

ence. In addition to light occupancy measures of fly activity, movement bout dynamics, and locomotor handedness were also recorded.

Temporal Light-Shade Assay

Ambient light preference is estimated in this assay by projecting a fully bright (intensity = 1) or fully dark (intensity = 0) stimulus to the entire arena when flies crossed an invisible virtual boundary in the center (min. power = 79µW±61nW, max. power = 2.4mW±4.7µW). Therefore, this stimulus paradigm was distinct from the spatial light-shade assay in that flies could not express a preference by navigating with respect to an apparent spatial pattern of light. A light occupancy measure was calculated for each fly as the fraction of time flies received the lit stimulus divided by the total duration of the experiment (2hrs). The invisible boundary zone was always positioned in the center of the arena such that a randomly exploring fly would receive both stimuli in roughly equal amounts. The angular orientation of the boundary zone was randomly initialized for each arena independently at the start of the experiment. To avoid rapid switching or flickering of the stimulus for flies sitting directly on the boundary, flies were required to cross a small buffer zone (5% of the arena diameter) beyond the arena center into the other zone before switching the stimulus.

Y-Maze Assay

In the Y-Maze assay, locomotor handedness was estimated as flies explored symmetrical Y-shaped arenas. Each time flies changed position from one arm of the maze to another turns were scored as left-handed or right-handed depending on whether the chosen arm was to the left of right on the choice point (i.e., the arena center). To avoid scoring centroid estimation errors around the choice point or small forays into new arms as turns, the flies were required to traverse a minimum distance (60% the length of the arm) into any arm before a turn was scored. In addition to locomotor handedness, measures of speed, movement bout dynamics, turn choice timing, and turn choice sequence were also recorded.

Central Complex Thermogenetic Screen

We extracted measure of turn timing clumpiness (choiceClumpiness) and turn direction switchiness (handSwitchiness) from a previously conducted thermogenetic screen for light dependent modulation of locomotor handedness (Kakaria et al., 2019). Gal4 driver and neuronal effector parental lines were crossed to generate F_1 progeny with sparse neuronal expression of effectors that could be thermogenetically activated (dTrpA1) or silenced ($Shibire^{ts}$). For all screen experiments, fly locomotor handedness was assayed for 4 hrs in the Y-Maze with the following temperature program: 1 h at 23°C (permissive temperature), 1 h temperature ramp up to the restrictive temperature ($Shibire^{ts}$ = 29°C, dTrpA1 = 32°C), 1 h at the restrictive temperature, and 1 h temperature ramp down from the restrictive temperature to 23°C. Although activity level varied over the duration of the experiment, flies made many turns throughout all temperature blocks: completing an average of 193±119 and 757±311 turns in the permissive and restrictive periods respectively. The background light was repeatedly switched on and off in time blocks (min = 5 seconds, other times = 10 and 30, max = 60 seconds , each hour had the same repeated sequence), in a random order shared between all screen experiments. Because we were interested only in temperature dependent modulation of turn dynamics, we computed individual clumpiness and switchiness scores for all turns within each temperature condition (permissive and restrictive) regardless of the light condition.

RNAseq and Transcriptomic Analyses

Single fly RNAseq preparation and sequencing was performed using TM3'seq, a high-throughput low-cost RNA protocol previously described in (Pallares et al., 2020). An overview of the method is provided below, but more detailed descriptions are available in an online protocol (*https://lufpa.github.io/TM3Seq-Pipeline/tm3seq_protocol*) and in the original publication.

Fly Tissue Preparation

Following Decathlon behavioral testing, flies were briefly anesthetized on CO_2 and transferred to a 96-well plate via aspiration. Immediately following, flies were flash frozen on liquid nitrogen and were decapitated to separate heads and bodies. Well plates were then transferred to a dry ice and ethanol cold bath to keep the samples frozen while heads were individually transferred to a separate 96-well plate with a cold probe needle. Samples were then stored at -80°C. After storage, tissue was ground in 100µl of lysis buffer with a 2.8mm stainless steel grinding bead steel grinding bead for 10 minutes at maximum speed with a homogenizer. CyBio® FeliX liquid handling robot (Analitik Jena) was used to perform mRNA extraction from the resulting lysate using a Dynabeads™ mRNA DIRECT™ Purification Kit (ThermoFisher, #61012) and a custom protocol (Pallares et al., 2020) optimized for low cost, yielding approximately 10 to 20 ng mRNA per head.

Library Preparation

RNA (10µl of 1ng/µl mRNA) was added to 1µl of 0.83uM oligo (Tn5ME-B-30T) for a 3 minute incubation at 65°C immediately prior to reverse transcription. The first strand of cDNA was synthesized by reverse transcription of the mRNA via a 1hr incubation at 42°C by adding the following to the reaction mixture above: 1µl SMARTScribe™ RT (Takara, #639538), 1µl dNTPs 10mM (NEB, #N0447S), 2µl DTT 0.1M (Takara, #639538), 4µl 5× First-Strand buffer (Takara, #639538), and 1µl B-tag-sw oligo. Following synthesis, the reverse transcriptase was inactivated via a 15 minute incubation 70°C. A cDNA amplification mixture was prepared by adding 5µl of the resulting first strand cDNA with 7.5µl of OneTaq HS Quick-load 2× (NEB, #M0486L) and 2.5µl water. The cDNA was then amplified in a thermocycler via the following program: 68°C 3min, 95°C 30sec, [95°C 10sec, 55°C 30sec, 68°C 3min] *3 cycles, 68°C 5min.

Tn5 tagmentation was used alongside universal adaptors for library amplification. The adapter-B was previously added during synthesis of the first cDNA strand. To create a Tn5 adaptor-A, an adapter annealing mixture was prepared by adding 10µl (100µM) of a forward oligo (adapter-A) and 10µl (100µM) reverse adapter-A oligo (Tn5MErev) to 80µl re-association buffer (10mM Tris pH 8.0, 50mM NaCl, 1mM EDTA). Oligos were annealed in a thermocycler with the following cycle program:

95°C for 10 minutes, 90°C for 1 minute followed by 60 cycles reducing temperature by 1°C/cycle, hold at 4°C. 5µl of 1µM annealed adapter was then anneal to 5µl of Tn5 in a thermal cycler for 30min at 37°C. 5µl of cDNA was mixed with 1µl of precharged Tn5. The adapter-A loaded Tn5 was diluted 7× in reassociation buffer: Glycerol (1:1)., 4µl of TAPS buffer 5× pH 8.5; 50mM TAPS, 25mM MgCl2, 50% v/v DMF), and 5µl of water. The mixture was then and incubated for 7min at 55°C followed by an additional 7 minute incubation with 3.5µl of SDS 0.2% (Promega, #V6551) to ensure that Tn5 was dissociated from the cDNA. The resulting cDNA libraries were then amplified. Briefly, 10µl of OneTaq HS Quick-Load 2x (NEB, #M0486L), 1µl i5 primer 1uM, 1µl i7 primer 1µM, and 7µl of water were used to amplify 1µl of the tagmentation reaction following the program: 68°C 3min, 95°C 30sec, [95°C 10sec, 55°C 30sec, 68°C 30sec] *12 cycles, 68°C 5min.

Sequencing and Expression Quantification

Samples were sequenced on an Illumina HiSeq 2500 in separate runs, using dual indexes and single-end ~100 bp sequencing. Low quality bases and adapter sequences were removed from reads using Trimmomatic 0.32 (SE ILLUMINACLIP:1:30:7 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:20; Bolger et al., 2014). Downstream analysis was only performed on reads at least 20 nucleotides after trimming. Reads were then mapped to the r6.14 *Drosophila melanogaster* genome assembly using STAR (Dobin et. al, 2012), and were assigned to genes using featureCounts from the package Subread (Liao et al., 2013). Read counts were further filtered to include only reads assigned to protein coding genes with a minimum of 500k reads per individual (n.b. measuring 3' biased expression).

Enrichment Analyses

Following sequencing, raw reads from individual flies with fewer than 1M reads total were removed. Raw read counts were normalized to reads per million (RPM) and were further filtered to exclude genes with less than 10 RPM averaged across all remaining flies. Individual fly read distributions were then quantile normalized so that the distributions of normalized reads were matched across all individual samples. One dimensional linear models were fit using expression data from the remaining 6,710 genes to predict each of the 97 individual behavioral scores (97x6,710 models total). Gene lists were constructed for each behavior separately, by collecting their respective significantly predictive genes (model p-value < 0.05).

Enrichment scores were computed using clusterProfiler (Yu et al., 2012) as part of the Bioconductor (www.bioconductor.org) package for R (version 3.6; R core team, 2020). We performed KEGG enrichment analysis on predictive gene lists from each behavior independently to identify functional categories of genes overrepresented for any particular behavioral measure compared to the background list of 6,710 genes. ClusterProfiler uses a Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) to adjust p-values. We report these adjusted p-values. KEGG categories were defined as significant if the adjusted p-values for the category was less than 0.05. Significant categories and their corresponding genes were identified separately for each behavior. Significant category (and gene) lists for groups of behaviors (i.e., all behaviors or a priori group behaviors) were constructed as the union of the significant enrichment categories (and genes) from each group's member behaviors. To assess reproducibility of enrichment categories and their corresponding gene sets, enrichment lists were repeatedly generated from bootstrap resampled individuals (500 replicates), downstream of quantile normalization. Average p-values were calculated as the mean across bootstrap replicates of the minimum adjusted p-value for each enrichment category across behaviors within a group. For each category, the average number of behaviors (Figures 4F, S11F) was calculated as the mean number of behaviors where the category was significant across bootstrap replicates.

Genomic Sequencing of Inbred Flies

DNA was extracted from individual flies using Zymo Quick-DNA Microprep Kit (Cat# D3020). Sequencing libraries were prepared using the transposase Tn5, using the same protocol described above (from cDNA). Samples were sequenced on an Illumina HiSeq 2500, using dual indexes, paired-end ~150 bp sequencing. After de-multiplexing, low quality bases and adapter sequences were removed from reads using Trimmomatic 0.32 (SE ILLUMINACLIP:1:30:7 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:20). Reads were then mapped to the r6.14 *Drosophila melanogaster* genome assembly using BWA -aln (Li et al., 2009). Genotypes were generated following GATK best practices (https://software.broadinstitute.org/gatk/best-practices/) (McKenna et al., 2010). Heterozygosity was estimated using VCFTools v0.1.16 --het (Danecek et al., 2011).

Merging Decathlon Data Sets

Inspection of raw measure distributions separated by batch (i.e. a single cohort of flies tested in a single behavioral box at the same time) showed batch effects on the sample means and dispersion even within the same Decathlon experiment. Therefore, data was z-scored separately by batch and measure type to control for sample differences in mean and variance. Data of the same assay, measure, and day of testing (i.e. the unique measures from a single day of Decathlon behavior) were then combined, resulting in an initial nxd data matrix for a single Decathlon experiment where n is the number of individuals and d is the number of measures. The matrix contained a substantial fraction of missingness (mean = 0.23±0.22) within measures after initial construction. We used alternating least squares (ALS) to estimate a complete matrix that preserved, as accurately as possible, the covariance structure of the underlying data (Figure S4). To reduce run to run variation in ALS, we generated 200 complete data matrices with ALS and computed a final complete median matrix of all repetitions. Simulations with ground truth data (Figure S4) with covariance structure similar to that of the Decathlon data sets showed that this process resulted in both lower error in the resulting matrices and lower run to run variation in the final matrix. This process resulted in three complete matrices: two nxd matrices for the two Berlin-K^{iso} Decathlon experiments and one nxd matrix for the NEX Decathlon experiment.

To combine the Decathlon-1 and Decathlon-2 matrices for Berlin-K^{iso}, we z-scored by behavioral measure within each matrix to adjust for Decathlon experiment batch effects on mean and variance. Data from the two matrices were then combined by matching unique assay and measure combination. Because

the order of assays was randomized for each Decathlon experiment, day of testing was ignored when combining behavioral measures from all non-circadian assays (e.g.,olfaction odorOccupancy from day 7 of Decathlon-1 was combined with olfaction odorOccupancy from day 8 of Decathlon-2). Circadian measures were matched by day of testing due to circadian measurements being collected on all days of testing (e.g.,circadian meanSpeed from day 1 of Decathlon-1 was combined with circadian meanSpeed from day 1 of Decathlon-2). Placeholder NaN values were inserted in cases where no matching measure existed in the other matrix (e.g.,temporal phototaxis for Decathlon-1). The resulting full data matrix was then z-scored by measure and any residual missing values (due to lack of an existing measure match) were then infilled via the ALS method described above.

Distilled Matrix Generation

We created a distilled matrix to condense any covarying features for which we had a prior expectation that the measures might be correlated for obvious or uninteresting reasons. We defined such groups of measures (a priori groups) which primarily consisted of either duplicate measures (e.g., all 10 days of circadian gravitaxis) or measures that are likely to be linked by the same underlying phenomenon (e.g., choiceNumber and meanSpeed across multiple assays). Details of the measures in each a priori group are detailed in supplemental table 2. We performed PCA on each a priori group separately in an attempt to capture group variance with fewer dimensions. We defined an adaptive cutoff for the number of principal components (PCs) retained from each group as the highest principal component above or within the 95% confidence interval of the variance explained by PCs computed on a shuffled version of the same matrix. We reasoned that PCs above or within the variance explained for PCs fit on the shuffled matrix (a matrix with approximately independent features) represented components fit to meaningful covariance of measures or components primarily aligned single independent measure.

Multidimensional Analyses

We estimated effective dimensionality of the Decathlon measures as the number of connected components in a thresholded covariance matrix represented as a directionless graph. We swept 200 threshold values uniformly spaced between the minimum and maximum covariance. A distribution of the effective dimensionality was formed by iteratively counting the number of connected components in the graph at each threshold value. To estimate a lower limit on the dimensionality and to assess the degree to which the number of connected components was dependent on specific groups of measures, we repeated the steps above on matrices after randomly selecting measures to drop from the matrix, iteratively increasing the number of measures dropped from 1 to d-1 (i.e., one feature remaining).

Unless otherwise specified, t-SNE embeddings of all data sets was performed on the euclidean pairwise distances of z-scored values with perplexity = 20. All embeddings were optimized by minimizing the KL-divergence between the original and embedded data. t-SNE performed on Decathlon measures was run with perplexity = 8 where we had an expectation that clusters would be relatively small (e.g., 3-8 data points) due to the low number of measures in our a priori groups (median = 5) and low number of unique measures from each assay (median = 8) and day of testing.

Unsupervised Behavioral Classification

As previously described (Berman et al., 2014), single fly videos were decomposed into behavioral classification time series with the motion mapper pipeline. Briefly, 200x200px frames centered on the flies were translationally and rotationally aligned with sub-pixel accuracy to a template fly body to restrict frame to frame variation to postural changes by the flies. The data dimensionality was reduced by restricting further analysis to the 6,700px with the highest variance. The data for all individuals was further compressed with PCA into 50 eigenmode postural time series. Principal component time series were then spectrally decomposed into 25 uniformly spaced frequency channels via Morlet wavelet transformation resulting in a high-dimensional (1,250) representation of each frame at various timescales. A representative sub-sampled training set of frames was constructed for each individual by embedding their data into two dimensions with t-SNE (van der Maatens and Hinton, 2008) and selecting frames according to proportionally to their local probability density in the embedded space. A joint two-dimensional embedding for all individuals data was then constructed by embedding the combined training set via t-SNE. Individual embeddings were generated by projecting all remaining data points into the joint embedding. As previously described, the distribution of log-speed trajectories within the embedded space was well-described by a two-component gaussian mixture model (GMM) with the majority of frames falling into a low speed mode. We used the standard deviation of the low speed GMM component to define a small gaussian kernel (σ=1.37) that was convolved with the embedded points the compute a continuous density. The embedded space was then segmented into discrete regions by computing watersheds (Najman and Schmitt, 1994) of the negative density. Behavioral classifications were then assigned to frames by the watershed region occupied. Frames were filtered by defining an embedded speed threshold (2.78) as the intersection of GMM components where the log-speed distribution was maximally separable. All frames above speed threshold received no classification. Individual occupancy within each classification was then used to compute a discrete probability density function for each individual.

Human readable labels were applied to each classification by generating 8x8 tiled movies of frames corresponding to individual pauses within each classification above a minimum duration of 10 ms. Labels were created by scoring which behaviors appeared to be frequently represented after repeated viewing. The labels were composed with the following format: body part (e.g., wing, forelimb, abdomen), the behaviors displayed (e.g., walking, grooming, movement), and a qualitative descriptor of the speed of movement (e.g., idle, slow, fast).

Statistics

Unless otherwise stated, all reported correlations were computed as the Spearman rank correlation. P-values reported for correlation coefficients were calculated via two-tailed t-test of the null hypothesis that the regression coefficient was not significantly different from zero.

To generate bootstrapped distributions of correlation the correlation matrices, Decathlon matrices (either the full or distilled matrices) were sorted to match measures across Decathlon data sets as described above. A pair of bootstrapped matrices were created by bootstrapping individuals from either the same (e.g., resampling Berlin-K^{iso} matrix twice) or different Decathlon matrices up to the size of the original matrix. The correlation matrix was computed for each bootstrapped matrix and the unique, off-diagonal *r*-values of each matrix were stored. A single correlation of correlations was then calculated on the two sets of *r*-values. The above steps were then repeated over 100 repetitions for all unique combinations of Decathlon data sets.

False Discovery Rate Calculation

We computed the false discovery rate (FDR) for significance of measure correlations as a function of α-value for each combination of Decathlon data set (Berlin-K^{iso} or NEX) and matrix type (full or distilled). Each data matrix was bootstrap resampled (100 repetitions) up to the size of the original matrix. Correlation matrices were computed for each matrix and *p*-values were calculated correlation coefficient (see statistics). We then calculated mean kernel density estimates for all *p*-values of unique measure combinations across bootstrap replicates. The above steps were then repeated for bootstrap shuffled data matrices to create a null distribution of *p*-values for correlation coefficients. We defined false discovery rate as the shuffled *p*-value density divided by the observed *p*-value density. For determining significance of correlation, we set $\alpha = 0.05$, corresponding to the following FDR: Berlin-K^{iso} full = 0.30, NEX full = 0.30, Berlin-K^{iso} distilled = 0.37, NEX distilled = 0.42.

Acknowledgements

We thank Ed Soucy, Brett Graham, Adam Bercu and Joel Greenwood of Harvard's CBS Neuroengineering core for help fabricating our instruments. Julie Peng helped with the genomic experiments. Gordon Berman was generous in consulting on the analysis and sharing data. Timothy Sackton provided insightful consultation and feedback on the RNAseq modeling and KEGG enrichment analysis. Joshua Shaevitz kindly shared code and expertise for the unsupervised analysis imaging rig. Tanya Wolff and Gerry Rubin kindly shared most of the Gal4 lines from the circuit screen. James Crall, Jennifer Erickson, Danylo Lavrentovich, and Shradda Lall provided helpful feedback on the manuscript. ZW and KSK were supported by NSF Graduate Research Fellowships DGE-1144152 and #2013170544; JFA was supported by the NIH under grant no. GM124881. BdB was supported by a Sloan Research Fellowship, a Klingenstein-Simons Fellowship Award, a Smith Family Odyssey Award, a Harvard/MIT Basic Neuroscience Grant, the NSF under grant no. IOS-1557913, and the NIH under grant no. MH119092.

Conflicts of Interest

The authors have no competing interests to declare.

References

Akhund-Zade, J., Ho, S., O'Leary, C., and Bivort, B. de (2019). The effect of environmental enrichment on behavioral variability depends on genotype, behavior, and type of enrichment. *Journal of Experimental Biology*.

Ayroles, J.F., Buchanan, S.M., Chelsea, O., Kyobi, S., Grenier, J.K., Clark, A.G., Hartl, D.L., and de Bivort, B.L. (2015). Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proc National Acad Sci 112*, 6706–6711.

Bartumeus, F., Catalan, J., Fulco, U., Lyra, M., and Viswanathan, G. (2002). Optimizing the encounter rate in biological interactions: Levy versus Brownian strategies. ´ *Phys. Rev. Lett. 88*, 097901.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological) 57*, 289–300.

Berdal, M.A. and Dochtermann, N.A. (2019). Adaptive Alignment of Plasticity With Genetic Variation and Selection. *Journal of Heredity 110*, 514–521.

Berman, G.J., Choi, D.M., Bialek, W., and Shaevitz, J.W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface 11*.

Bierbach, D., Laskowski, K., and Wolf, M. (2017). Behavioural individuality in clonal fish arises despite near-identical rearing conditions. *Nature communications 8*.

Bolger, A., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. Bioinformatics (Oxford, England) 30, 2114–2120.

Bruijning, M., Metcalf, C.J.E., Jongejans, E., and Ayroles, J.F. (2020). The Evolution of Variance Control. *Trends in Ecology and Evolution 35*, 22–23.

Buchanan, S.M., Kain, J.S., and de Bivort, B.L. (2015). Neuronal control of locomotor handedness in *Drosophila*. *Proc National Acad Sci 112*, 6700–6705.

Calhoun, A., Pillow, J., and Murthy, M. (2019). Unsupervised identification of the internal states that shape natural behavior. *Nature Neuroscience 22*, 1–10.

Cande, J., Namiki, S., Qiu, J., Korff, W., Card, G.M., Shaevitz, J.W., Stern, D.L., and Berman, G.J. (2018). Optogenetic dissection of descending behavioral control in *Drosophila*. *eLife 7*, e34275.

Charmantier, A., Garant, D., and Kruuk, L. (2014). Quantitative genetics in the wild (Oxford: Oxford University Press). ISBN: 978-0-19-967424-4.

Claridge-Chang, A., Roorda, R., Vrontou, E., Sjulson, L., Li, H., Hirsh, J., and Miesenböck, ¨ G. (2009). Writing memories with light-addressable reinforcement circuitry. *Cell 139*, 405–415.

Danecek, P. et al. (2011). The variant call format and VCFtools. *Bioinformatics 27*, 2156– 2158.

Dobin, A., Davis, C., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson.

M., and Gingeras, T. (2012). STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England) 29, 15–21.

Dochtermann, N.A., Schwab, T., Anderson Berdal, M., Dalos, J., and Royaute, R. (2019). The Heritability of Behavior: A Meta-analysis. *J. Hered. 110*, 403–410.

Freund, J., Brandmaier, A.M., Lewejohann, L., Kirste, I., Kritzler, M., Krüger, A., Sachser, N., Lindenberger, U., and Kempermann, G. (2013). Emergence of Individuality in Genetically Identical Mice. *Science 340*, 756–759.

Haddad, S. and Marder, E. (2018). Circuit Robustness to Temperature Perturbation Is Altered by Neuromodulators. *Neuron 100*, 609–623.e3.

Hamada, F., Rosenzweig, M., Kang, K., Pulver, S.R., Ghezzi, A., Jegla, T.J., and Garrity, P. (2008). An internal thermal sensor controlling temperature preference in *Drosophila*. *Nature 454*, 217–220.

Honegger, K. and Bivort, B. de (2018). Stochasticity, individuality and behavior. *Current Biology 28*, R8–R12.

Honegger, K.S., Smith, M., Churgin, M.A., Turner, G.C., and de Bivort, B.L. (2019). Idiosyncratic neural coding and neuromodulation of olfactory individuality in *Drosophila*. *Proceedings of the National Academy of Sciences*.

Hopper, K. (1999). Risk-spreading and bet-hedging in insect population biology. *Annu. Rev. Entomol. 44*, 535–60.

Huntingford, F. (1976). The relationship between anti-predator behaviour and aggression among conspecifics in the three-spined stickleback, *Gasterosteus aculeatus*. *Animal Behaviour 24*, 245–260.

Jeanson, R. and Weidenmuller, A. (2014). Interindividual variability in social insects - proximate causes and ultimate consequences. *Biological Reviews 89*, 671–687.

Jenett, A., Rubin, G., Ngo, T., Shepherd, D., and Murphy, C. (2012). A GAL4-driver line resource for *Drosophila* neurobiology. *Cell reports 2*, 991–1001.

Johnson, J.C. and Sih, A. (2007). Fear, food, sex and parental care: a syndrome of boldness in the fishing spider, *Dolomedes triton*. *Animal Behaviour* 74, 1131–1138.

Kain, J.S., Stokes, C., and de Bivort, B.L. (2012). Phototactic personality in fruit flies and its suppression by serotonin and white. *Proc National Acad Sci 109*, 19834–19839.

Kain, J.S., Zhang, S., Jamilla, A., Samuel, A.D., Klein, M., and de Bivort, B.L. (2015). Variability in thermal and phototactic preferences in *Drosophila* may reflect an adaptive bet-hedging strategy. *Evolution 69*, 3171–3185.

Kain, J., Stokes, C., Gaudry, Q., Song, X., Foley, J., Wilson, R., and de Bivort, B.L. (2013). Leg-tracking and automated behavioural classification in *Drosophila*. *Nature Communications 4*.

Kakaria, K. and de Bivort, B.L., (2017). Ring attractor dynamics emerge from a spiking model of the entire protocerebral bridge. *Frontiers in behavioral neuroscience 11*.

Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research 28*, 27–30.

Kitamoto, T. (2001). Conditional modification of behavior in *Drosophila* by targeted expression of a temperature-sensitive shibire allele in defined neurons. *Journal of Neurobiology 47*, 81–92.

Kottler, B., Faville, R., Bridi, J., and Hirth, F. (2019). Inverse Control of Turning Behavior by Dopamine D1 Receptor Signaling in Columnar and Ring Neurons of the Central Complex in *Drosophila*. *Current Biology 29*, 567–577.

Krakauer, J., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., and Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron 93*, 480–490.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics 25*, 1754–1760.

Liao, Y., Smyth, G., and Shi, W. (2013). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics (Oxford, England) 30, 923–930.

Lin, Y., Chen, Z.-X., Oliver, B., and Harbison, S.T. (2016). Microenvironmental Gene Expression Plasticity Among Individual *Drosophila melanogaster*. *G3: Genes-Genomes Genetics 6*, 4197–4210.

Linneweber, G.A. et al. (2020). A neurodevelopmental origin of behavioral individuality in the *Drosophila* visual system. *Science 367*, 1112–1119.

Long, A., Macdonald, S., and King, E. (2014). Dissecting Complex Traits Using the *Drosophila* Synthetic Population Resource. *Trends in genetics 30*, 488–495.

Lynch, M. and Walsh, B. (1998). Genetics and analysis of quantitative traits 1st ed. (Sunderland, Mass: Sinauer). ISBN: 0878934812.

Maaten, L. and Hinton, G. (2008). Visualizing data using *t*-SNE. *Journal of machine learning research 9*, 2579–2605.

Mackay, T.F., et al. (2012). The *Drosophila melanogaster* Genetic Reference Panel. *Nature 482*.

Mann, K., Deny, S., Ganguli, S., and Clandinin, T.R. (2020). Causal coupling between neural activity, metabolism, and behavior across the *Drosophila* brain. *bioRxiv*.

Matthews, G., Deary, I., and Whiteman, M. (2003). Personality traits 2nd ed. (Cambridge University Press).

Mckenna, A. et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research 20*, 1297-1303.

Mellert, D.J., Williamson, W.R., Shirangi, T.R., Card, G.M., and Truman, J.W. (2016). Genetic and Environmental Control of Neurodevelopmental Robustness in *Drosophila*. *PLOS ONE 11*, 1–29.

Mootha, V. et al. (2003). PGC-1α−responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics 34*, 267–273.

Najman, L. and Schmitt, M. (1994). Watershed of a continuous function. *Signal Processing 38*, 99–112.

Namiki, S., Dickinson, M.H., Wong, A.M., Korff, W., and Card, G.M. (2018). The functional organization of descending sensory-motor pathways in *Drosophila*. *eLife 7*, e34272.

Nöthel, H. (1981). Investigations on radiosensitive and radioresistant populations of *Drosophila melanogaster* VIII. The system of relative radioresistance in immature oocytes of the irradiated population ROI4. *Mutation Research*.

Sakai, O. (2018). Comparison of personality between juveniles and adults in clonal gecko species. *Journal of ethology 36*, 221–228.

Oers, K. van, Drent, P.J., Goede, P. de, and Noordwijk, A.J. van (2004). Realized heritability and repeatability of risk-taking behaviour in relation to avian personalities. *Proceedings of the Royal Society of London. Series B: Biological Sciences 271*, 65–73.

Ofstad, T., Zuker, C., and Reiser, M. (2011). Visual place learning in *Drosophila melanogaster*. *Nature 474*, 204–207.

Pallares, L.F., Picard, S., and Ayroles, J.F. (2020). TM3'seq: A Tagmentation-Mediated 3' Sequencing Approach for Improving Scalability of RNAseq Experiments. *G3: Genes, Genomes, Genetics 10*, 143–150.

Pantoja, C., Hoagland, A., Carroll, E., and Karalis, V. (2016). Neuromodulatory regulation of behavioral individuality in zebrafish. *Neuron 91*, 587–601.

Goldberg, .L. (1993). The structure of phenotypic personality traits. *American psychologist*.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. (Vienna, Austria).

Reale, D., Reader, S., Sol, D., McDougall, P., and Dingemanse, N. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews of the Cambridge Philosophical Society 82*, 291–318.

Robie, A. et al. (2017). Mapping the Neural Substrates of Behavior. *Cell 170*, 393–406.e28.

Schuett, W., Dall, S., Baeumer, J., Kloesener, M.H., Nakagawa, S., Beinlich, F., and Eggers, T. (2011). Personality variation in a clonal insect: The pea aphid, *Acyrthosiphon pisum*. *Dev Psychobiol 53*, 631–640.

Seeds, A.M., Ravbar, P., Chung, P., Hampel, S., Midgley Frank M, J., Mensh, B.D., and Simpson, J.H. (2014). A suppression hierarchy among competing motor programs drives sequential grooming in *Drosophila*. *eLife 3*. Ed. by R.L. Calabrese, e02951.

Seelig, J. and Jayaraman, V. (2015). Neural dynamics for landmark orientation and angular path integration. *Nature 521*, 186–191.

Shah, K., Hirokawa, J., Magani, P., and Maimon, G. (2017). A neural circuit architecture for angular integration in *Drosophila*. *Nature 546*.

Skutt-Kakaria, K., Reimers, P., Currier, T.A., Werkhoven, Z., and de Bivort, B.L. (2019). A neural circuit basis for context-modulation of individual locomotor behavior. *bioRxiv*.

Stern, S., Kirst, C., and Bargmann, C. (2017). Neuromodulatory control of long-term behavioral patterns and individuality across development. *Cell 171*, 1649–1662.

Todd, J.G., Kain, J.S., and de Bivort, B.L. (2017). Systematic exploration of unsupervised methods for mapping behavior. *Phys Biol 14*, 015002.

Werkhoven, Z., Rohrsen, C., Qin, C., Brembs, B., and de Bivort, B.L. (2019). MARGO (Massively Automated Real-time GUI for Object-tracking), a platform for high-throughput ethology. *PLOS ONE 14*, 1–24.

Wolff, T. and Rubin, G.M. (2018). Neuroarchitecture of the *Drosophila* central complex: A catalog of nodulus and asymmetrical body neurons and a revision of the protocerebral bridge catalog. *Journal of Comparative Neurology 526*, 2585–2611.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology 16*, 284–287.

Supplementary materials

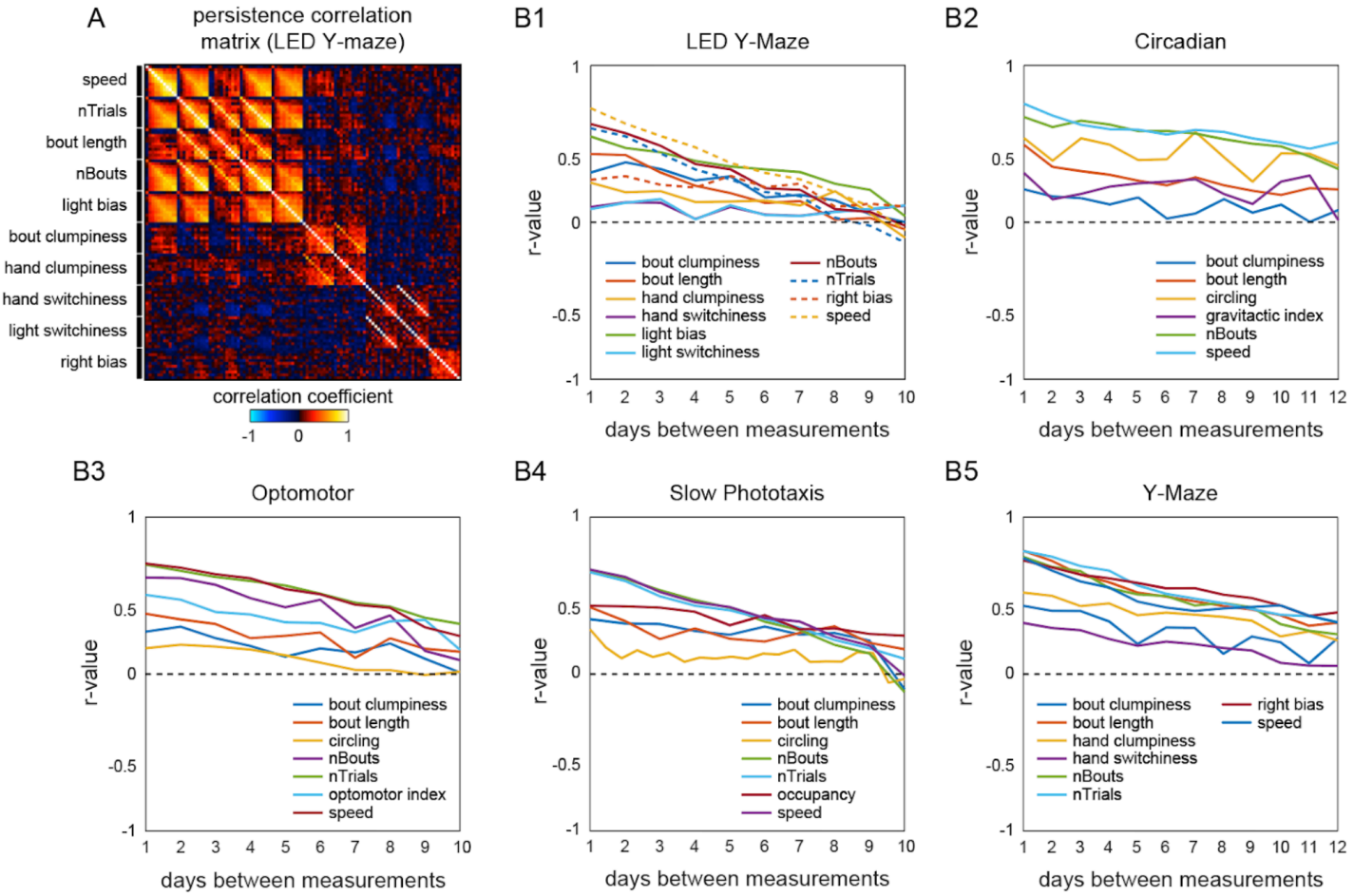


Figure S1 — Persistence of primary behavioral measures across assays. A) Example correlation matrix for 10 retests of the same individuals on a single assay (phototactic Y-Maze) across subsequent days. Rows and columns were sorted to cluster the same measures across days. B1-5) Plots of average correlation coefficient (*r*-value) as a function of the days between measurements. Colors indicate the measure within each plot separately (see legends). Dashed black line denotes no correlation (r=0). Measures in all assays showed persistent individual variation that diminished over time.

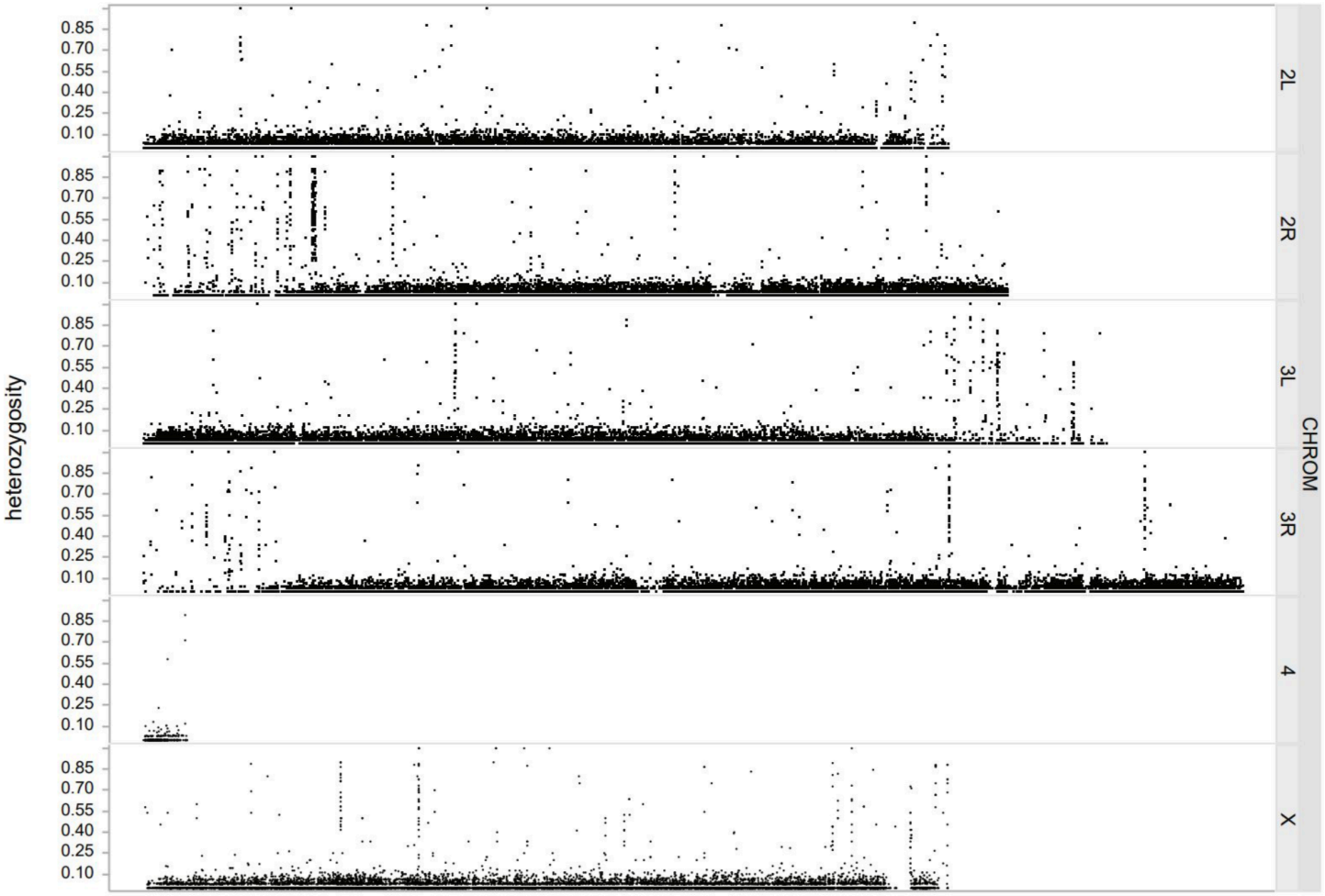


Figure S2 — Genomic sequencing to confirm isogeny of Berlin-K^{iso}. Fraction of Berlin-K^{iso} flies (n=48) that were heterozygous at any given position in the genome. Typically no more than 10% of flies appeared to be variant at a given locus. There was evidence of residual heterozygosity at approximately 75 sites throughout the genome. These can be seen as distinct regions appearing as "columns" of heterozygous sites, and are most found at the ends of chromosome arms.

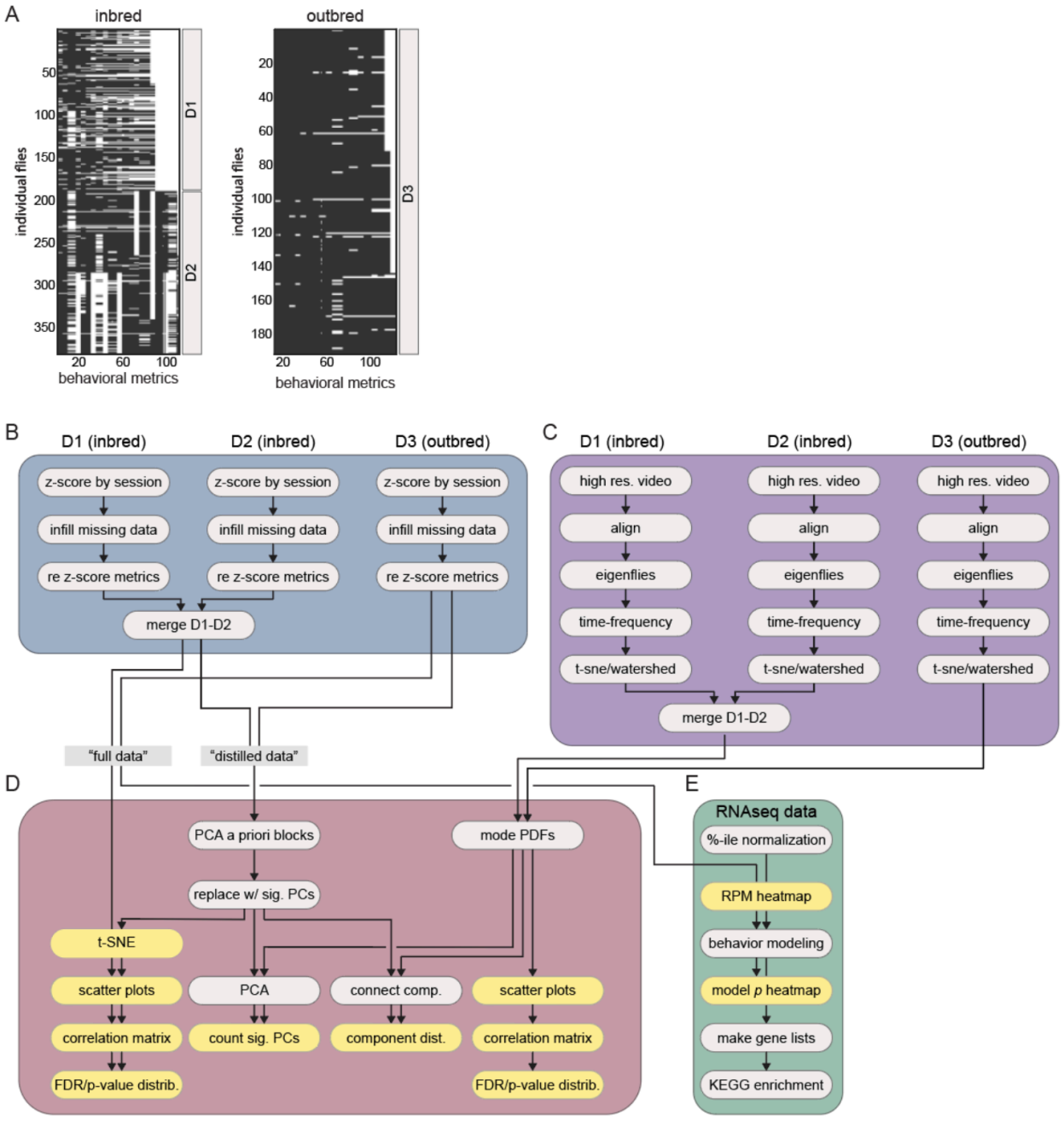


Figure S3 — Schematic of the Decathlon analysis pipeline. A) Data matrices from the inbred and outbred Decathlon experiments. Rows are flies, columns are behavioral measures. White cells indicate missing data prior to infilling. B) Decathlon assay behavioral data preprocessing pipeline. Measures are *z*-scored by imaging session to adjust for batch effects. Missing data is infilled with an ALS imputed matrix calculated as the median of 200 ALS imputations. C) Data acquisition and analysis for the unsupervised behavioral classification pipeline. High resolution, high frame rate single fly videos are aligned to a template fly and compressed into principal component time series (i.e., eigenflies). PC time series are then decomposed into 25 frequency spectral time series via Morlet wavelet transformation. The resulting high dimensional data is then embedded into two dimensions with *t*-SNE and then clustered into discrete behavioral modes via watershed transformation. D) Diagram of Decathlon behavioral data analysis. Data from the Decathlon assays and unsupervised behavioral classifications are analyzed separately. Inbred decathlon batches D1 and D2 merged into a single behavioral data matrix before analyzing inbred (left) and outbred (right) flies separately. The covariance structure and effective dimensionality of the resulting matrices are then analyzed independently or are compressed into a "distilled" matrix with fewer dimensions. Distilled matrices are generated by retaining significant PCs within each a priori measure group with variance explained higher than or within the variance explained by a model of *n*-independent dimensions (see methods). E). Individual flies undergo gene expression profiling via RNAseq. Behavioral metrics are modeled independently by gene expression. Gene lists are then constructed for each behavior by thresholding model significance. Lists are then used as input for KEGG pathway enrichment analysis.

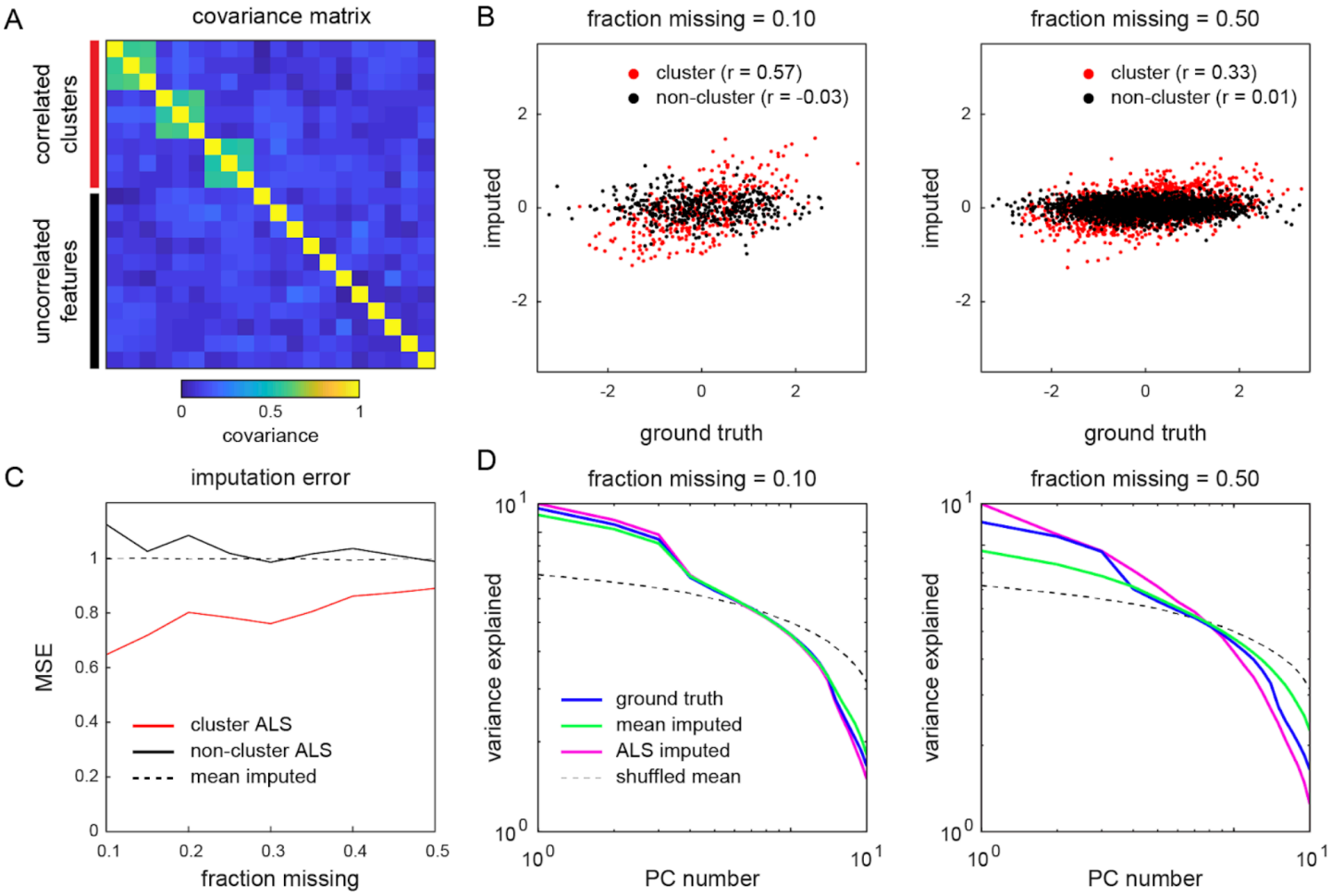


Figure S4 — Toy data comparison of matrix-infilling methods. A) Covariance matrix of a toy data set generated from a multivariate normal distribution with covariance = 0.5 for features in correlated clusters (red block) and covariance = 0 for independent features. B) Scatter plots of data from a ground truth vs the same data infilled with average ALS imputation (200 repetitions) after randomly deleting either 10% (left) and 50% (right) of the entries. Color indicates whether the data belonged to a correlated cluster as in A. C) Comparison of mean squared error (MSE) for mean infilled and ALS average infilled data. D) Log scree plots of the variance explained for ranked principal components resulting from PCA on the ground truth, mean infilled, average ALS infilled, and shuffled matrix.

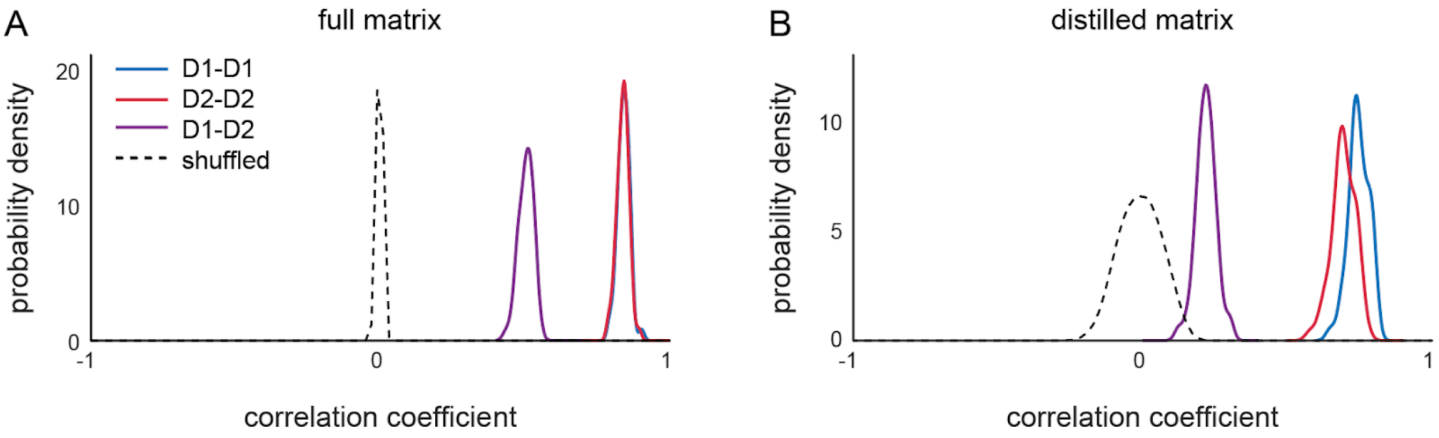


Figure S5 — Correlation of correlation matrix values between D1 vs D2 and shuffled matrices. A) Distributions of correlation of full matrix correlation matrices across bootstrap replicates. Correlation coefficients were calculated by bootstrap resampling data matrices (Decathlon-1 or Decathlon-2) and computing the correlation matrix. Correlation was then computed between two matrices (i.e., either the same or different matrices across resamplings). B) Distributions of correlation of distilled matrix correlation matrices across bootstrap replicates as in A.

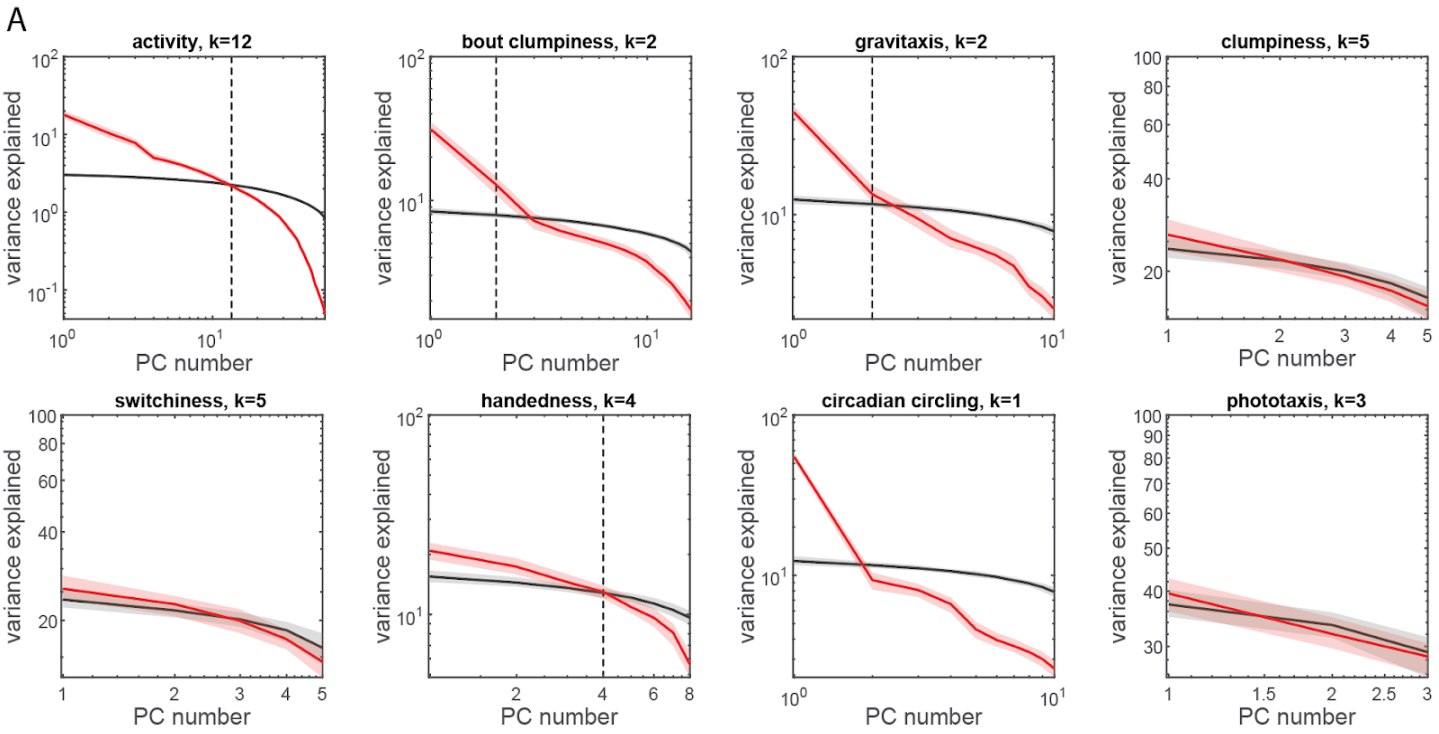
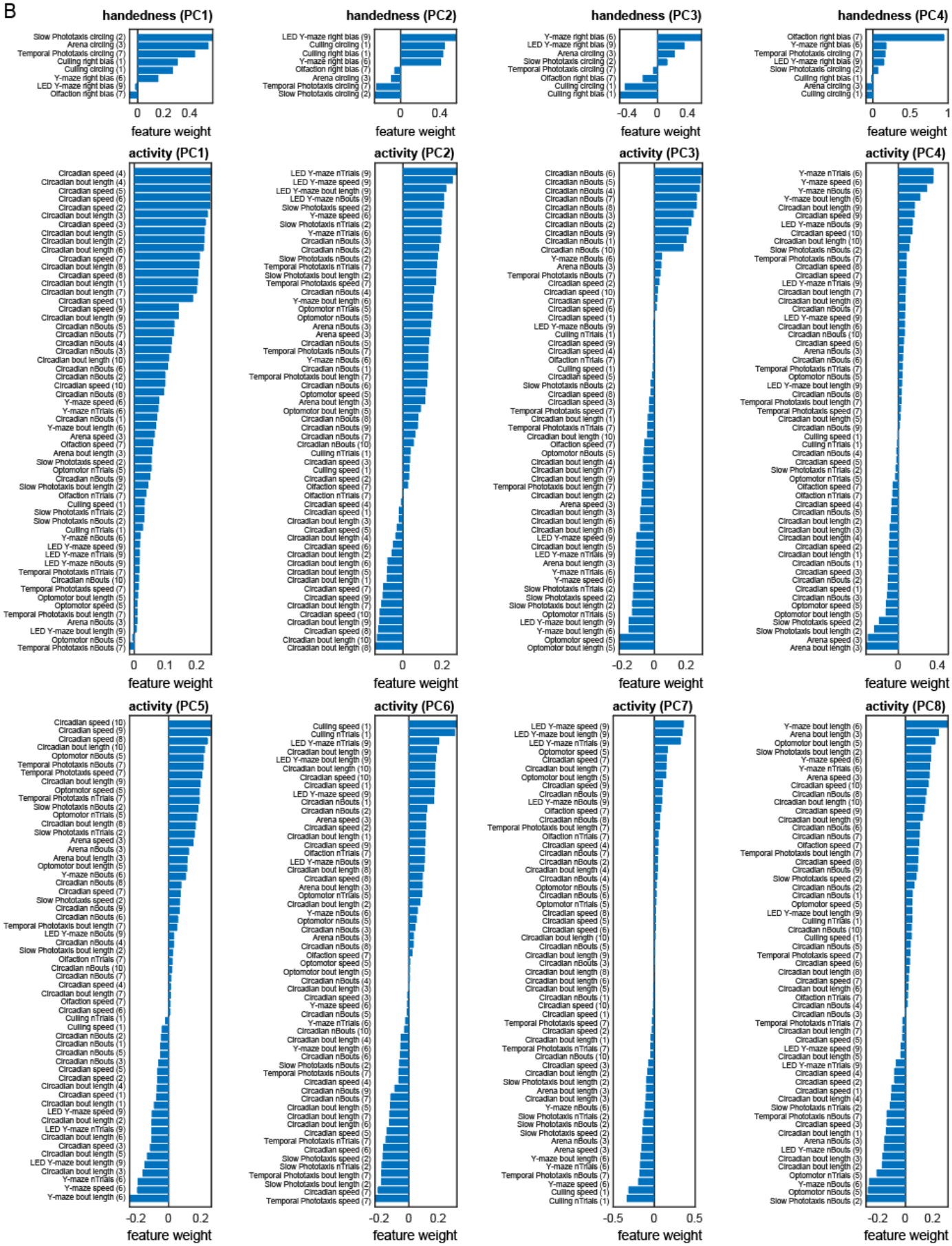
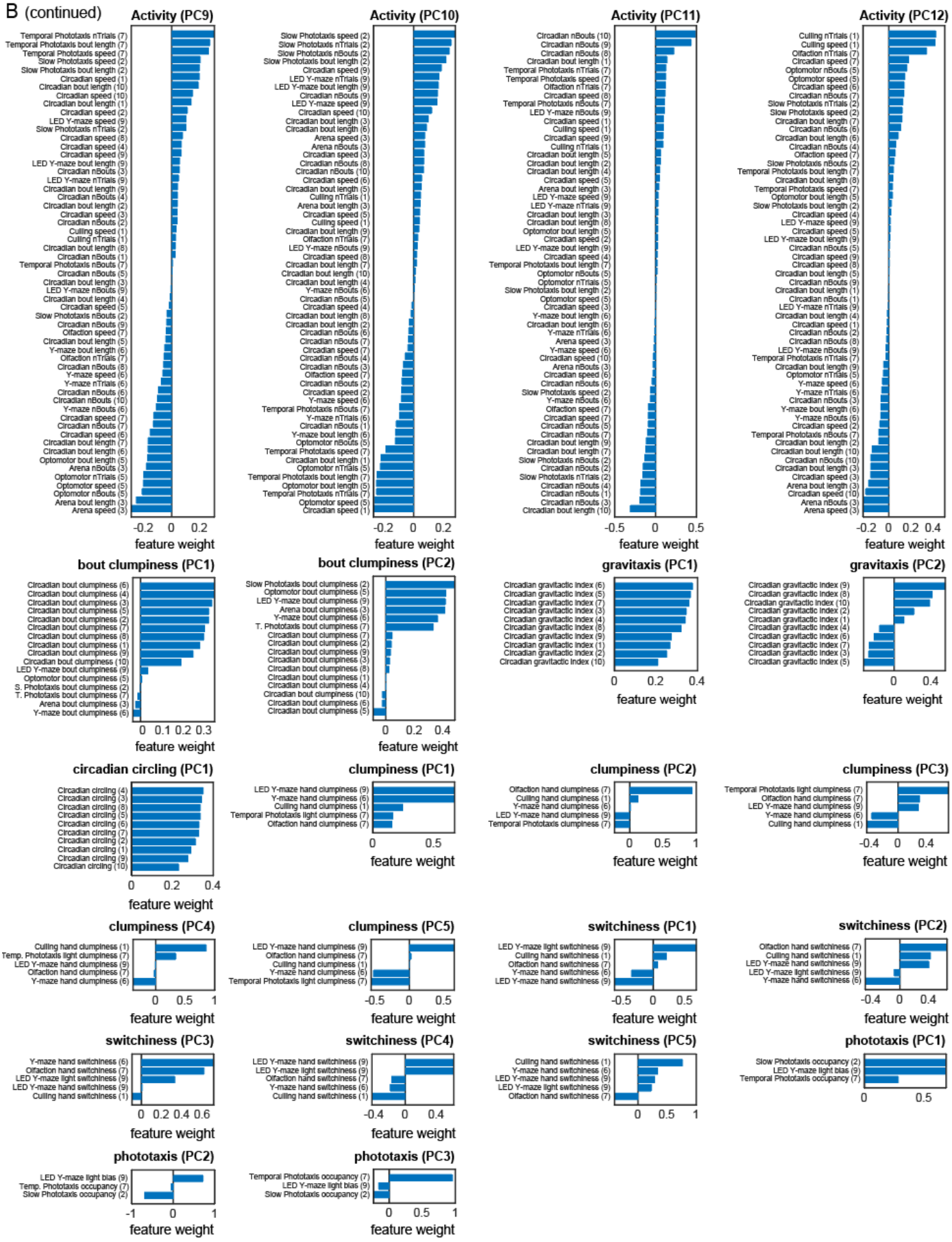


Figure S6 (continues on next two pages)— PCA of a priori groups. A) Log scree plots of the normalized ranked eigenvalues (i.e., variance explained) for PCA performed on measures from each a priori group separately. Color indicates observed (red) or shuffled (black) data. Shaded regions correspond to 95% confidence intervals as calculated by bootstrap resampling. The number of significant PCs (*k*) was calculated as the highest rank PC above or within the 95% confidence interval of the shuffled matrix variance explained. B) Measure loadings of significant PCs for each a priori group.

B



B (continued)



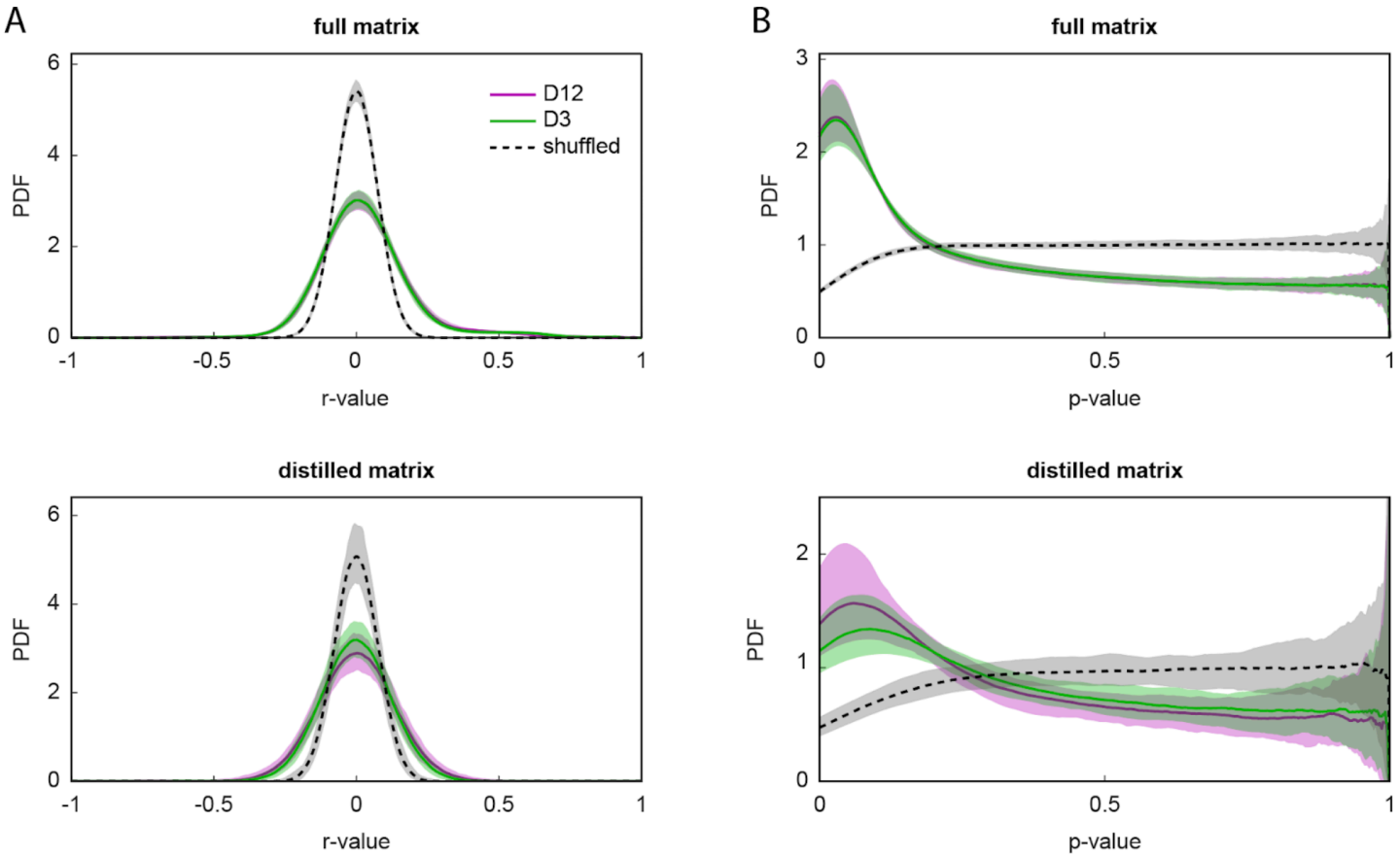


Figure S7 — Distribution of correlation coefficients and *p*-values in the full and distilled correlation matrix. A) Kernel density estimates of the unique (i.e., lower matrix triangle) correlation coefficients in the full (top) and distilled (bottom) correlation matrices. Distributions exclude duplicate pairwise and self correlations. B) Kernel density of the unique *p*-values for the correlation coefficients in A. In all plots, dashed lines indicate distributions for column-wise (i.e., within each feature) shuffled matrices. Shaded regions correspond to 95% confidence intervals calculated by bootstrap resampling.

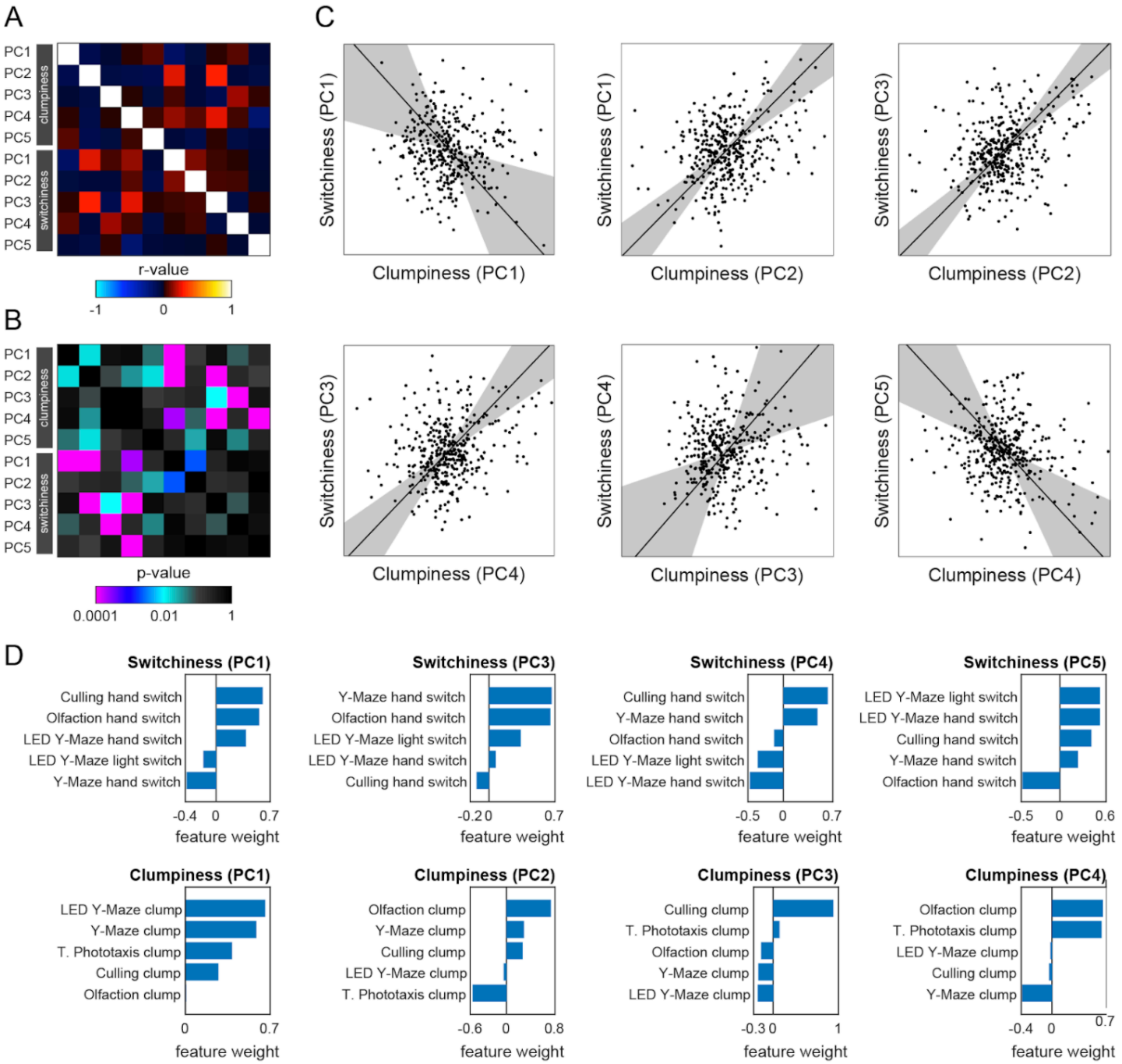


Figure S8 — Significant correlations among the principal components of switchiness and clumpiness. A) Subset of the distilled correlation matrix corresponding to the significant PCs of switchiness and clumpiness. B) *P*-value matrix for the correlation coefficients in A. C) Scatter plots of significant correlations between switchiness and clumpiness. Points correspond to individual flies. Line indicates the line of best fit and the shaded region indicates the 95% confidence interval of the fit as calculated by bootstrap resampling. D) Measure loadings for the PCs in C.

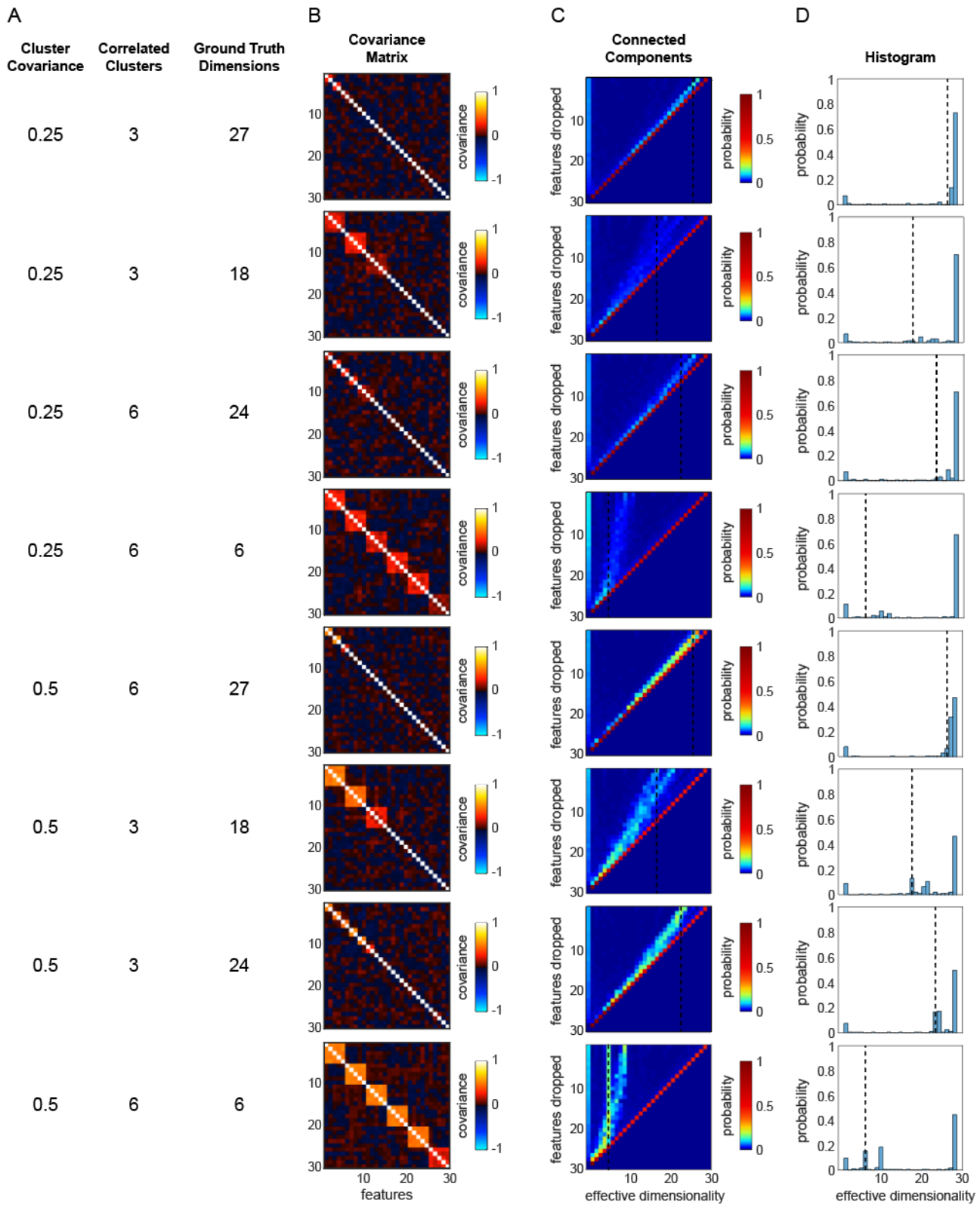


Figure S9 — Effective dimensionality simulations with toy data. A) Parameters used to generate a 200x30 ground truth matrix from a multivariate normal distribution as in Figure S4. Cluster covariance refers to the covariance between features belonging to the same correlated cluster. Correlated clusters corresponds to the total number of such clusters. Ground truth dimensionality is the sum of number of correlated clusters and the total number of independent features. B) Covariance matrices of toy data sets with correlated clusters along the diagonal. One cluster in each data set has covariance equal to half of the remaining clusters. C) Effective dimensions heatmap as a function of number of features retained in the toy data after dropping *n* random features. Rows correspond to histograms of number of connected components in the covariance matrices. D) Connected components histogram for the full covariance matrix (i.e., the top row of C).

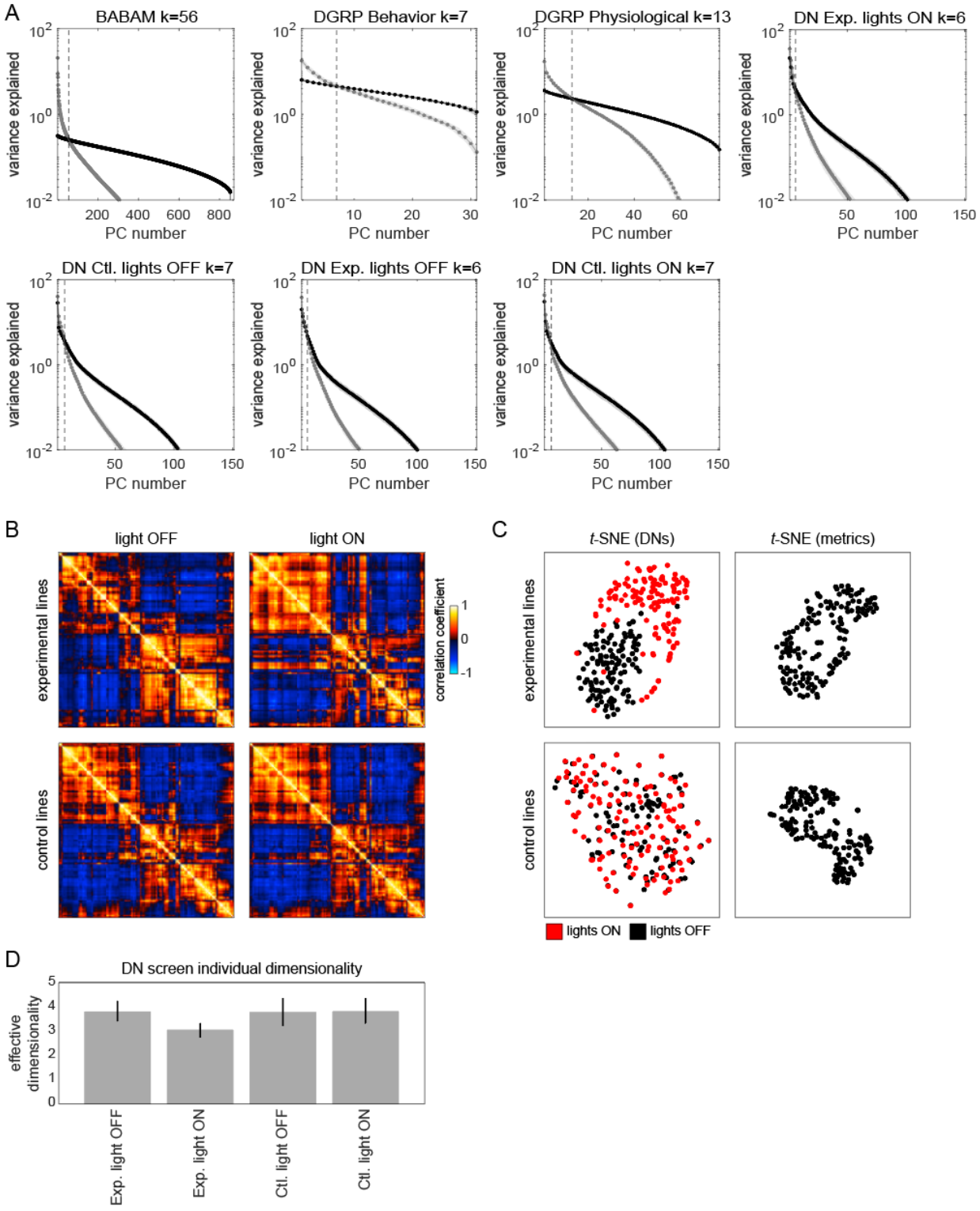


Figure S10 — Structure of behavioral variation in non-Decathlon data sets. A) Scree plots showing the variance explained for each principal component of the BABAM Gal4 screen, DGRP (behavioral and physiological), and descending neuron screen (all experimental groups and conditions) behavioral data sets. Point colors indicate variance explained for the observed (gray) and shuffled (black) data matrices. The dashed line denotes one measure of effective dimensionality, where the variance explained of the observed principal components is below the 95% confidence interval (shaded regions) of the shuffled data. B) Correlation matrices for the combined behavioral PDF for each descending neuron set separated by experimental group and condition. C) *t*-SNE embeddings of the descending neurons lines (left) and unsupervised measures (right) from the descending neuron screen. Color of the data points in the left hand plot indicates whether individuals were control (black; Gal4/+) or experimental (red; Gal4/UAS-CsChrimson). D) Average effective dimensionality (as measured by the intersection of observed and shuffled ranked PC variances) of the individual behavioral PDFs separated by experimental group and condition. Error bars are the 95% confidence interval of the mean.

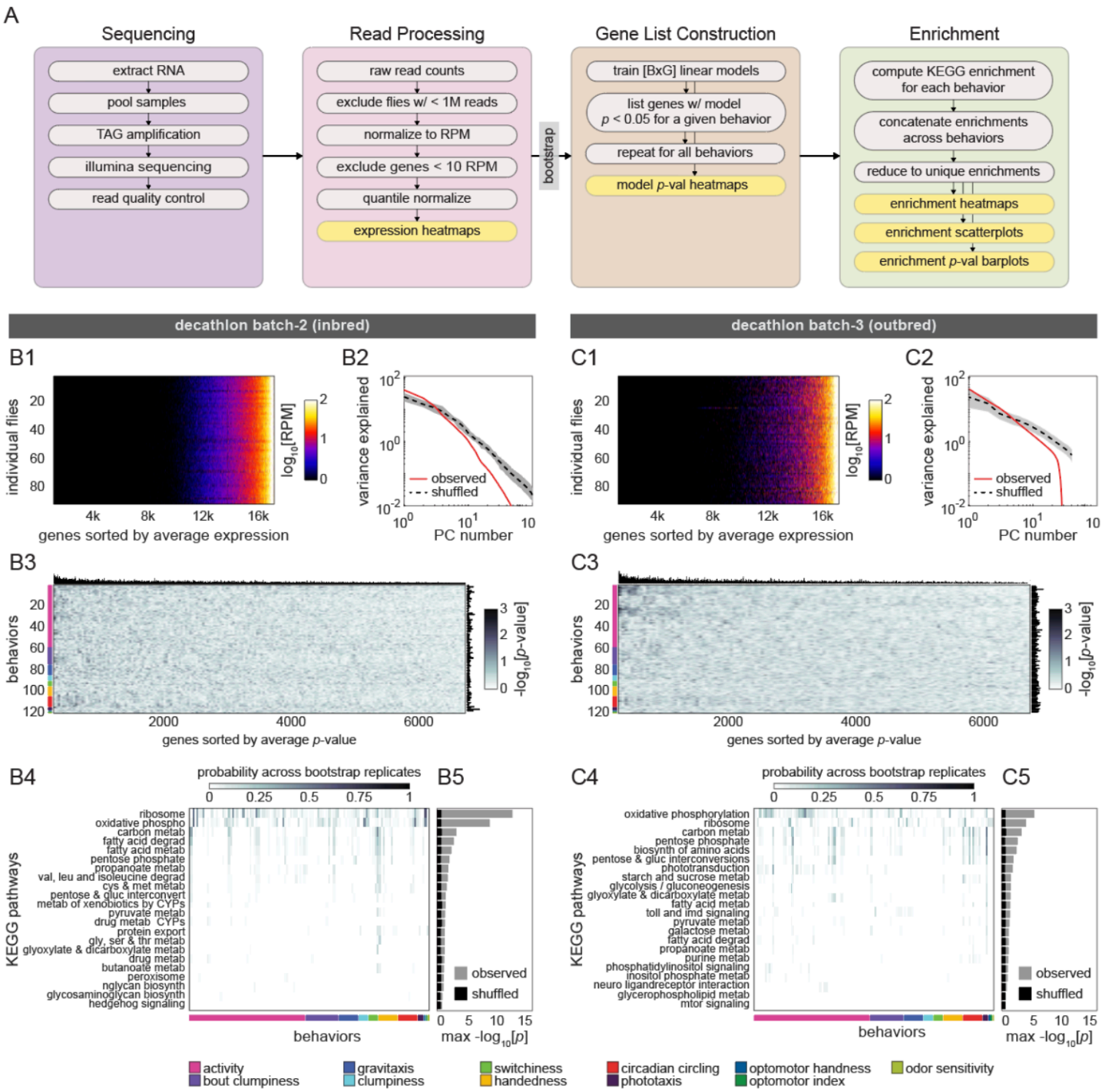


Figure S11 — Correlation between individual transcriptomes and behavioral biases. A) Schematic of the RNAseq analysis pipeline for both experimental batches. Following sequencing, raw read counts were filtered and quantile normalized. Linear models were fit using individual expression of single genes [G] to predict individual behavioral scores [B] (BxG models total). Gene lists were constructed for each behavior separately from its set of significantly predictive genes. To create a comprehensive list of significantly enriched (adjusted $p < 0.05$) KEGG categories for all behaviors, the union of all behaviors' significant gene lists was determined (see methods). B-C) Gene expression characterization, modeling, and KEGG pathway enrichment for the second inbred (B1-B5) and outbred (C1-C5) Decathlon experiments. B1-C1) Individual log reads per million (RPM) from showing the complete unfiltered list of 17,470 genes sorted by mean expression level. B2-C2) Scree plots showing the logged % variance explained for the first 100 PCs from PCA performed on observed (red) and shuffled (black) gene expression data. Shaded region corresponds to 95% CI as calculated by bootstrap resampling. B3-C3) Heatmaps of linear model -log *p*-values. Colored bars (left) indicate a priori group identity of behavioral measures (rows). Bar graphs show the number of significant ($p<0.05$) models for each gene (top) and behavior (right). B4-C4) Heatmaps showing the probability across bootstrap replicates of a KEGG pathway being significantly enriched in the list of predictive genes for a given behavior. B5-C5) Bar plot showing the average across bootstrap replicates of the maximum (across behaviors) negative log adjusted *p*-value of all enriched KEGG pathways. Color indicates results from observed (gray) or shuffled control (black) data.

Supplementary Movies

Movie S1 — *https://www.youtube.com/watch?v=3cuSapTnSOU* — Examples of a mode of walking behavior as identified by the unsupervised analysis, from movies of single flies, made up of successive frames classified as the same behavior. Colored dots indicate whether flies are outbred (NEX; red) or inbred (Berlin-K^{iso}; blue).

Movie S2 — *https://www.youtube.com/watch?v=E1vIl-1nLZU* — Examples of a mode of wing grooming behavior as identified by the unsupervised analysis, from movies of single flies, made up of successive frames classified as the same behavior. Colored dots indicate whether flies are outbred (NEX; red) or inbred (Berlin-K^{iso}; blue).

Movie S3 — *https://www.youtube.com/watch?v=PYC4NONOv7s* — Examples of a mode of head grooming behavior as identified by the unsupervised analysis, from movies of single flies, made up of successive frames classified as the same behavior. Colored dots indicate whether flies are outbred (NEX; red) or inbred (Berlin-K^{iso}; blue).

Movie S4 — *https://www.youtube.com/watch?v=lwCetEi0siA* — Examples of a mode of abdomen flexing behavior as identified by the unsupervised analysis, from movies of single flies, made up of successive frames classified as the same behavior. Colored dots indicate whether flies are outbred (NEX; red) or inbred (Berlin-K^{iso}; blue).

Supplementary Tables

Name	Description	Metrics Collected	Arena Geometry	Throughput	Duration	Activity Inclusion Criteria
Activity Screening	Activity measurement profiling at the start of the decathlon experiment used to narrow future testing from a cohort of 288 flies to the most active 192 flies	circling bias, turn bias, choice number, turn clumpiness, hand switchiness, mean speed, bout number, bout length, bout clumpiness	Y-shaped arena with arms equally spaced apart at 120 degrees (12.7mm length for each maze arm)	96 flies per behavioral box	2 hrs	highest 192 choice number
Arena Circling	Open exploration of circular arena.	circling bias, mean speed, bout number, bout length, bout clumpiness	circular (28mm diameter)	48 flies per behavioral box	2 hrs	mean speed > centroid error
Circadian	Overnight activity level monitoring on food.	circling bias, mean speed, bout number, bout length, bout clumpiness, gravitactic index	96-multiwell plate (6.8mm diameter, 10.5mm depth)	96 flies per behavioral box	20-21 hrs	mean speed > centroid error
LED Y-maze Phototaxis	Discrete two-choice (light or no light) phototactic assay. Choices are scored as turns toward or away from a lit arm of the maze and as turns to the right or left of the choice point.	circling bias, turn bias, photoBias, choice number, turn clumpiness, hand switchiness, photo switchiness, mean speed, bout number, bout length, bout clumpiness	Y-shaped arena with LEDs at the end of each arm (12.7mm length for each maze arm)	72 flies per behavioral box	2 hrs	choice number > 10
Olfactory Sensitivity	Odor vs no-odor adaptation assay measuring sensitivity to a low concentration (typically aversive) odorant	odor occupancy, preodor occupancy, turn bias, choice clumpiness, hand switchiness, choice number, mean speed	linear tunnels with odorant ports at either end and decision boundary in center	15 flies per experiment	15 min (3 min preodor, 12 min odor)	mean speed > centroid error
Optomotor	Optomotor response assay where flies are repeatedly targeted with a high-contrast rotating pinwheel stimulus projected onto the arena floor	circling bias, opto index, mean speed, trial number, bout number, bout length, bout clumpiness	circular (28mm diameter)	48 flies per behavioral box	2 hrs	mean speed > centroid error
Spatial Shade-Light	Phototaxis assay where dark and light occupancy are measured. Individual arenas are targeted with a half lit, half dark spatial pattern.	circling bias, mean speed, bout number, bout length, bout clumpiness, photo occupancy	circular (28mm diameter)	48 flies per behavioral box	2 hrs	mean speed > centroid error
Temporal Shade-Light	Phototaxis assay where dark and light occupancy are measured. Individual arenas are either fully dark or fully lit depending on fly position relative to an invisible virtual boundary.	circling bias, mean speed, bout number, bout length, bout clumpiness, photo occupancy	circular (28mm diameter)	48 flies per behavioral box	2 hrs	mean speed > centroid error
Y-maze	Discrete two-choice (right or left) locomotor handedness assay. Choices are scored as turns to the right or left of the choice point in a Y-shaped arena.	circling bias, turn bias, choice number, turn clumpiness, hand switchiness, mean speed, bout number, bout length, bout clumpiness	Y-shaped arena with arms equally spaced apart at 120 degrees (12.7mm length for each maze arm)	96 flies per behavioral box	2 hrs	choice number > 10

Table S1 — Details of Decathlon behavioral assays.

Name	Metrics	Metric Number	PCs Retained	% Variance Captured
Activity	Circadian: bout length (1-10), bout number (1-10), mean speed (1-10) Arena Circling: bout length, bout number, mean speed Culling: trial number, mean speed LED Y-Maze: bout length, bout number, mean speed, trial number Olfaction: trial number, mean speed Optomotor: bout length, bout number, trial number, mean speed Shade-Light: bout length, bout number, trial number, mean speed Temporal Shade-Light: bout length, bout number, trial number, mean speed Y-Maze: bout length, bout number, trial number, mean speed	57	13	68.36
Bout Clumpiness	Circadian: bout clumpiness (1-10) Arena Circling: bout clumpiness Culling: bout clumpiness LED Y-Maze: bout clumpiness Optomotor: bout clumpiness Shade-Light: bout clumpiness Temporal Shade-Light: bout clumpiness Y-Maze: bout clumpiness	16	2	43.55
Gravitaxis	Circadian: gravitaxis (1-10)	10	2	57.72
Clumpiness	Culling: turn clumpiness LED Y-Maze: turn clumpiness Olfaction: turn clumpiness Temporal Shade-Light: choice clumpiness Y-Maze: turn clumpiness	5	5	100
Switchiness	Culling: hand switchiness LED Y-Maze: hand switchiness, light switchiness Olfaction: hand switchiness Y-Maze: hand switchiness	5	5	100
Handedness	Arena Circling: circling Culling: right bias LED Y-Maze: right bias Olfaction: right bias Shade-Light: circling Temporal Shade-Light: circling Y-Maze: right bias	8	4	65.02
Circadian Handedness	Circadian: circling (1-10)	10	1	55.11
Optomotor Handedness	Optomotor: circling	1	1	100
Phototaxis	LED Y-Maze: light bias Shade-Light: light occupancy Temporal Shade-Light: light occupancy	3	3	100
Optomotor Index	Optomotor: index	1	1	100
Odor Sensitivity	Olfaction: odor occupancy	1	1	100

Table S2 — Breakdown of a priori groups (group name, measures in each group, number of measures in each group, number of PCs kept from each group, variance explained by kept PCs).

Targeted Neurons	Gal4 Lines	Source
P-FN3	SS00079, SS00406, SS02255, SS02221, SS00007, SS02304, SS00083	Gift from Tanya Wolff and Gerry Rubin
P-FN4	SS00161, SS00078, SS00019, SS00082, SS00083	Gift from Tanya Wolff and Gerry Rubin
P-FN5	SS02243, SS02209, SS02211, SS00082	Gift from Tanya Wolff and Gerry Rubin
P-EN	SS02241, SS02268, SS02232	Gift from Tanya Wolff and Gerry Rubin
P-EG	SS00097	Gift from Tanya Wolff and Gerry Rubin
E-PG	SS00090, SS00131, SS02258	Gift from Tanya Wolff and Gerry Rubin
E-PG9	SS04781, SS04782, SS02374	Gift from Tanya Wolff and Gerry Rubin
P-F-gall	SS02267, SS02271	Gift from Tanya Wolff and Gerry Rubin
P-F-ROB	SS02192, SS02293	Gift from Tanya Wolff and Gerry Rubin
PF-CreL	SS02252, SS02214	Gift from Tanya Wolff and Gerry Rubin
PsL-PB	SS02216, SS02200	Gift from Tanya Wolff and Gerry Rubin
PBCap	SS04775	Gift from Tanya Wolff and Gerry Rubin
PBInt	SS02231, SS02235, SS00116	Gift from Tanya Wolff and Gerry Rubin
Ps-P	SS02296, SS00155, SS00153	Gift from Tanya Wolff and Gerry Rubin
PsIb-P	SS04769	Gift from Tanya Wolff and Gerry Rubin

Table S3 — List of Gal4 lines and the cell types they target from the Shibire^{ts} and dTRPA1 screen.

Decathlon-1 (inbred)							
KEGG Pathway ID	Description	Class	Sub-class	Observed p-value	Avg. min. p-value	Gene hits	Predictive genes
dme00190	Oxidative phosphorylation	metabolism	energy	3.04E-04	9.6E-14	17	ATPsyndelta, ATPsynE, ATPsynF, ATPsynO, CG5037, COX5B, cype, levy, ND-18, ND-19, ND-20, ND-23, ND-24, ND-39, ND-ACP, ND-B16.6, ND-MWFE, Vha14-1, Vha16-1, Vha26, Vha55, Vha68-2, VhaAC45, VhaM9.7-b, VhaSFD
dme03010	Ribosome	genetic information processing	translation	1.47E-07	1.4E-12	36	mRpS10, mRpS17, RpL10, RpL10Ab, RpL11, RpL12, RpL13, RpL13A, RpL17, RpL18, RpL18A, RpL24-like, RpL3, RpL32, RpL34a, RpL35A, RpL37a, RpL39, RpL4, RpL5, RpL7, RpL9, RpLP0, RpLP1, RpLP2, RpS13, RpS16, RpS19a, RpS23, RpS25, RpS3A, RpS4, RpS5a, RpS7, RpS9, sta
dme03050	Proteasome	genetic information processing	folding, sorting, and degradation	5.61E-15	4.0E-09	27	Pl31, Pomp, Prosalpha2, Prosalpha3, Prosalpha4, Prosalpha5, Prosalpha6, Prosalpha7, Prosbeta1, Prosbeta2, Prosbeta3, Prosbeta4, Prosbeta5, Prosbeta7, REG, Rpn10, Rpn11, Rpn13, Rpn2, Rpn3, Rpn5, Rpn7, Rpn8, Rpt1, Rpt3, Rpt6, Sem1
dme01200	Carbon metabolism	metabolism	global	6.13E-05	5.2E-05	30	aay, AcCoAS, Acox57D-p, Ald1, Arc42, Cat, CG10932, CG17333, CG4390, CG5214, CG5577, CG6287, CG7430, CG8036, CG9527, Echs1, Eno, fbp, Fdh, Gapdh1, Gapdh2, Gdh, Got1, Irp-1B, Mdh1, Mdh2, Men, Men-b, Mtpalpha, muc, Nc73EF, PCB, Pdha, Pdhb, Pfk, Pgi, Pglym78, ppl, Scsalpha1, ScsbetaA, Shmt, Spat, Taldo, Tpi
dme00071	Fatty acid degradation	metabolism	lipids	4.89E-05	1.6E-04	12	Acox57D-p, Acsl, Aldh, Arc42, bgm, CG10932, CG31075, CG3902, CG7461, CG9527, Echs1, Fdh, Mtpalpha, Mtpbeta, yip2
dme01212	Fatty acid metabolism	metabolism	lipid	2.07E-04	4.0E-04	14	Acox57D-p, Acsl, Arc42, bgm, CG10932, CG3902, CG7461, CG9527, Desat1, Echs1, FASN1, Mtpalpha, Mtpbeta, Ppt2, yip2
dme04624	Toll and Imd signaling pathway	organismal systems	immune system	1.14E-03	1.1E-03	13	BomS1, BomS2, BomS3, cact, dl, hep, Jra, nec, PGRP-LA, PGRP-SA, Rel, Tak1, Uev1A
dme04745	Phototransduction - fly	organismal systems	sensory system	2.93E-02	3.1E-03	8	Act5C, Eip63F-1, Galphaq, Gbeta76C, inaC, inaD, ninaC, norpA, Plc21C, trp
dme00010	Glycolysis / Gluconeogenesis	metabolism	carbohydrate	2.55E-03	3.6E-03	14	AcCoAS, Ald1, Aldh, Aldh-III, CG31075, CG32444, CG7430, CG9008, Eno, fbp, Fdh, Gapdh1, Gapdh2, muc, Pdha, Pdhb, Pepck2, Pfk, Pgi, Pglym78, Pgm1, Tpi
dme00280	Valine, leucine and isoleucine degradation	metabolism	amino acid	1.17E-03	4.4E-03	10	Aldh, Arc42, CG10932, CG11241, CG1673, CG31075, CG3902, CG7430, Echs1, Gabat, Hmgs, Mtpalpha, Mtpbeta, scu, yip2
dme01230	Biosynthesis of amino acids	metabolism	amino acid	4.04E-04	5.0E-03	18	aay, Ald1, Argl, Cbs, CG10184, CG1673, CG6287, CG8036, Eip55E, Eno, Gapdh1, Gapdh2, Got1, Gs1, Gs2, Hn, Irp-1B, P5CS, PCB, Pfk, Pglym78, Sam-S, Shmt, Taldo, Tpi
dme00270	Cysteine and methionine metabolism	metabolism	amino acid	3.84E-04	5.1E-03	14	Adi1, Ahcy, AhcyL1, Cbs, CG11241, CG1673, CG6287, Eip55E, Gclc, Gclm, Got1, Kyat, Mdh1, Mdh2, Mtap, Sam-S, SamDC
dme00640	Propanoate metabolism	metabolism	carbohydrate	4.24E-04	5.1E-03	10	AcCoAS, Acox57D-p, Arc42, CG10932, CG7430, CG9527, Echs1, Gabat, Mtpalpha, Scsalpha1, ScsbetaA
dme00620	Pyruvate metabolism	metabolism	carbohydrate	1.26E-03	5.7E-03	9	AcCoAS, Aldh, CG10932, CG31075, CG7430, Mdh1, Mdh2, Men, Men-b, muc, PCB, Pdha, Pdhb, Pepck2, tzn

dme00630	Glyoxylate and dicarboxylate metabolism	metabolism	carbohydrate	3.82E-03	7.9E-03	10	AcCoAS, Cat, CG10932, CG5577, CG7430, Gs1, Gs2, Irp-1B, Mdh1, Mdh2, ppl, Shmt, Spat
dme00250	Alanine, aspartate and glutamate metabolism	metabolism	amino acid	2.26E-03	1.4E-02	11	Argl, CG11241, CG8132, Gabat, Gad1, Gdh, Got1, Gs1, Gs2, P5CDh1, Prat2, Spat
dme00983	Drug metabolism - other enzymes	metabolism	xenobiotics degradation	1.38E-02	1.5E-02	13	CG15117, CG6330, CG8360, GstD2, GstE1, GstE12, GstE13, GstE2, GstE3, GstE6, GstO1, l(2)k01209, pyd3, Ugt316A1, Ugt36E1, Ugt36F1, Ugt49B1
dme04144	Endocytosis	cellular processes	transport and catabolism	1.20E-02	1.8E-02	21	Arf102F, ArfGAP1, Arpc4, Arpc5, CenB1A, Chc, cpb, Egfr, EndoA, Gprk2, Ist1, Lerp, Rab11, Rab5, Sec71, shi, Smox, Src64B, Vps36, Vps37B, Vps4, Vps60, WASp
dme00980	Metabolism of xenobiotics by cytochrome P450	metabolism	xenobiotics degradation	2.57E-02	2.3E-02	8	Aldh-III, Fdh, GstD2, GstE1, GstE12, GstE13, GstE2, GstE3, GstE6, GstO1, Ugt316A1, Ugt36E1, Ugt36F1, Ugt49B1
dme04310	Wnt signaling pathway	environmental information processing	signal transduction	3.42E-02	2.4E-02	15	Axn, CanA-14F, CkIalpha, CkIIalpha, dco, dlp, fz, inaC, Jra, nej, norpA, pk, Pka-C1, Plc21C, RhoL, Roc1a, sgg, Smox, Tak1, Vang
dme00982	Drug metabolism - cytochrome P450	metabolism	xenobiotics degradation	2.57E-02	2.4E-02	8	Aldh-III, Fdh, GstD2, GstE1, GstE12, GstE13, GstE2, GstE3, GstE6, GstO1, Ugt316A1, Ugt36E1, Ugt36F1, Ugt49B1
dme03008	Ribosome biogenesis in eukaryotes	genetic information processing	translation	9.39E-04	2.5E-02	14	Bka, CG1671, CG30349, CG4866, CkIIalpha, eIF6, Fib, NHP2, Non1, Nop56, Nop60B, Ns1, Ns2, pcm, Ran, Rat1
dme04145	Phagosome	cellular processes	transport and catabolism	1.88E-02	2.5E-02	17	Act5C, betaTub97EF, CG11459, Cp1, Dlic, Lamp1, Rab5, RhoL, Sec61alpha, Sec61beta, Sec61gamma, Syx18, Vha14-1, Vha16-1, Vha26, Vha55, Vha68-2, VhaAC45, VhaM9.7-b, VhaSFD
dme00260	Glycine, serine and threonine metabolism	metabolism	amino acid	3.78E-03	2.6E-02	9	aay, Cbs, CG10184, CG10361, CG11241, CG6287, CG7430, Eip55E, Pglym78, ppl, Shmt, Spat
dme04142	Lysosome	cellular processes	transport and catabolism	1.07E-02	3.6E-02	12	cathD, CG11459, CG12163, CG15117, CG3376, CG4847, Chc, Cp1, Hexo1, Ids, Lamp1, Lerp, Ppt2, Sap-r, Tsp42Ed, Vha16-1, VhaAC45, VhaSFD
dme00650	Butanoate metabolism	metabolism	carbohydrate metabolism	7.91E-03	3.7E-02	7	Arc42, CG10932, Echs1, Gabat, Gad1, Hmgs, L2HGDH, Mtpalpha
dme00052	Galactose metabolism	metabolism	carbohydrate	4.47E-02	3.8E-02	6	CG32444, CG9436, Gale, Galt, Pfk, Pgm1, UGP
dme00970	Aminoacyl-tRNA biosynthesis	genetic information processing	translation	3.40E-02	3.8E-02	9	AlaRS, ArgRS, CysRS, GlnRS, GlyRS, MetRS, SerRS, ThrRS, TyrRS
dme00330	Arginine and proline metabolism	metabolism	amino acid metabolism	9.55E-03	4.0E-02	8	Aldh, Argk, CG31075, Got1, Oat, Odc1, P5CDh1, P5CS, SamDC, slgA
dme00500	Starch and sucrose metabolism	metabolism	carbohydrate	2.51E-02	4.3E-02	7	AGBE, CG5171, GlyP, Gyg, Pgi, Pgm1, Tps1, UGP
dme00380	Tryptophan metabolism	metabolism	amino acid	3.41E-02	4.5E-02	6	Aldh, Cat, CG10932, CG31075, CG5214, CG7430, Echs1, Kyat, Mtpalpha, v
dme00981	Insect hormone biosynthesis	metabolism	terpenoids and polyketides	4.44E-03	6.2E-02	5	Aldh, antdh, CG15739, CG31075, CG40485, Jheh2
dme00600	Sphingolipid metabolism	metabolism	lipid metabolism	1.07E-02	8.3E-02	7	bwa, Cerk, CG10425, CG3376, ifc, laza, wun
dme03060	Protein export	genetic information processing	folding, sorting, and degradation	2.70E-02	8.5E-02	7	CG11110, Sec61alpha, Sec61beta, Sec61gamma, Spase25, Srp19, Srp9, SrpRalpha
dme04341	Hedgehog signaling pathway - fly	environmental information processing	signal transduction	9.15E-03	1.2E-01	8	CkIalpha, dco, gish, Gprk2, Pka-C1, Roc1a, sgg, Su(fu)
dme04130	SNARE interactions in vesicular transport	genetic information processing	folding, sorting, and degradation	1.06E-02	1.2E-01	7	Bet1, Gos28, Syb, Syx18, Syx8, Vti1b, Ykt6
dme03018	RNA degradation	genetic information processing	folding, sorting, and degradation	2.13E-02	1.8E-01	8	CG2021, DCP1, Eno, PAN3, pcm, Pfk, Rat1, Rga, Tob
dme00860	Porphyrin and chlorophyll metabolism	metabolism	cofactors and vitamins	4.90E-02	2.2E-01	6	Cchl, CG15117, CG5037, Ppox, Ugt316A1, Ugt36E1, Ugt36F1, Ugt49B1

Decathlon-2 (inbred)

KEGG Pathway ID	Description	Class	Sub-class	Observed p-value	Avg. min. p-value	Gene hits	Predictive genes
dme03010	Ribosome	genetic information processing	translation	4.2E-09	7.95E-11	31	mRpL12, mRpS17, RpL12, RpL13, RpL13A, RpL17, RpL18, RpL24, RpL28, RpL34b, RpL35, RpL37a, RpL37A, RpL7A, RpL8, RpL9, RpLP1, RpS15, RpS16, RpS17, RpS18, RpS19a, RpS2, RpS23, RpS25, RpS26, RpS29, RpS30, RpS4, sta

dme00190	Oxidative phosphorylation	metabolism	energy	3.3E-05	4.03E-08	22	ATPsynD, CG9065, COX4, COX5A, COX5B, cype, ND-19, ND-51, ND-75, ND-ASHI, ND-B14, ND-B14.5A, ND-B17.2, ND-PDSW, ox, RFeSP, SdhA, SdhB, SdhC, SdhD, UQCR-C2, UQCR-Q, Vha36-1
dme01200	Carbon metabolism	metabolism	global	6.0E-04	3.7E-03	17	Acox57D-p, Arc42, CG1640, CG5214, CG6287, CG8036, CG9527, fbp, Gapdh1, Hex-A, Mdh1, Men-b, PCB, ppl, ScsbetaA, SdhA, SdhB, SdhC, SdhD, Spat, Taldo
dme00071	Fatty acid degradation	metabolism	lipids	6.0E-04	9.7E-03	9	Acox57D-p, Arc42, CG3902, CG7461, CG9527, CPT2, Mtpbeta, yip2
dme01212	Fatty acid metabolism	metabolism	lipid	1.9E-02	1.7E-02	9	Acox57D-p, Arc42, CG3902, CG7461, CG9527, CPT2, Mtpbeta, yip2
dme00980	Metabolism of xenobiotics by cytochrome P450	metabolism	xenobiotics degradation	1.3E-02	6.9E-02	6	Acsl, GstD1, GstD2, GstT3, Ugt301D1, Ugt49B1
dme00982	Drug metabolism - cytochrome P450	metabolism	xenobiotics degradation	1.3E-02	8.2E-02	6	Acsl, GstD1, GstD2, GstT3, Ugt301D1, Ugt49B1
dme00983	Drug metabolism - other enzymes	metabolism	xenobiotics degradation	1.3E-02	9.4E-02	7	Acsl, GstD1, GstD2, GstE3, GstT3, Ugt301D1, Ugt49B1
dme00350	Tyrosine metabolism	metabolism	amino acid	2.0E-02	1E-01	5	Faa, GstZ2, Hpd, l(2)k01209
dme03060	Protein export	genetic information processing	folding, sorting, and degradation	3.2E-02	1.3E-01	4	hgo, Sec61beta, Sec61gamma, Trp1
dme00510	N-Glycan biosynthesis	metabolism	glycan biosynthesis	1.7E-02	1.6E-01	7	Alg2, Alg3, Alg7, Dpm1, OstDelta, twr, xit
dme04144	Endocytosis	cellular processes	transport and catabolism	2.4E-02	1.7E-01	14	ALiX, aPKC, Cbl, CG2224, CG33303, CG5498, Eps-15, par-6, PIP4K, Smurf, Stam, Usp8, Vps45
dme00534	Glycosaminoglycan biosynthesis - heparan sulfate / heparin	metabolism	glycan biosynthesis	3.8E-02	1.9E-01	5	AP-2mu, botv, Hs3st-B, Hs6st, Hsepi
dme00052	Galactose metabolism	metabolism	carbohydrate	4.3E-02	2.1E-01	4	CG10863, CG6083, GlcAT-I, Hex-A
dme00051	Fructose and mannose metabolism	metabolism	carbohydrate	4.3E-02	2.9E-01	4	CG10863, CG6083, fbp, GlcAT-I, Hex-A
dme00790	Folate biosynthesis	metabolism	cofactors and vitamins	4.3E-02	3.1E-01	5	Akr1B, CG10863, CG6083, GlcAT-I, se

Decathlon-3 (outbred)

KEGG Pathway ID	Description	Class	Sub-class	Observed p-value	Avg. min. p-value	Gene hits	Predictive genes
dme00190	Oxidative phosphorylation	metabolism	energy	4.31E-04	8.1E-06	23	ATPsynB, ATPsynCF6, ATPsynF, ATPsynG, COX7C, COX8, cype, ND-15, ND-19, ND-B14, ND-B14.5A, ND-B14.7, ND-B16.6, ND-B17, ND-PDSW, SdhC, sun, UQCR-6.4, Vha14-1, Vha44, Vha55, Vha68-1, VhaPPA1-1
dme03010	Ribosome	genetic information processing	translation	9.34E-04	1.60E-04	19	mRpL2, mRpL22, mRpS17, mRpS9, RpL13, RpL15, RpL22, RpL26, RpL29, RpL32, RpL36A, RpL38, RpS20, RpS21, RpS25, RpS26, RpS27, RpS3, RpS30
dme01200	Carbon metabolism	metabolism	global	1.58E-05	8.23E-04	24	AcCoAS, ACOX1, CG6287, CG7430, CG8036, CG9527, fbp, Got2, Irp-1B, Mdh1, Men, Mtpalpha, PCB, Pgd, Pgi, Pglym78, Prps, Rpe, Rpi, SdhC, Shmt, Spat, Taldo, Zw
dme00030	Pentose phosphate	metabolism	carbohydrate	5.06E-04	3.24E-03	10	CG8036, fbp, Pgd, Pgi, Pgm1, Prps, Rpe, Rpi, Taldo, Zw
dme01230	Biosynthesis of amino acids	metabolism	amino acid	4.31E-03	4.99E-03	13	CG6287, CG8036, Got2, Hn, Irp-1B, PCB, Pglym78, Prps, Rpe, Rpi, Sam-S, Shmt, Taldo
dme00040	Pentose and glycuronate interconversions	metabolism	carbohydrate	6.03E-03	1.60E-02	10	Akr1B, CG3534, CG3597, CG3609, CG7322, Rpe, sgl, Sodh-1, UGP, Ugt35B1
dme04745	Phototransduction - fly	organismal systems	sensory system	3.91E-03	2.0E-02	10	Cam, Gbeta76C, Itpr, ninaC, ninaE, norpA, Pkc53E, Plc21C, rdgC, trp
dme00500	Starch and sucrose metabolism	metabolism	carbohydrate	6.95E-03	3.9E-02	7	AGBE, CG5171, GlyP, GlyS, Pgi, Pgm1, UGP
dme00010	Glycolysis / Gluconeogenesis	metabolism	carbohydrate	2.31E-02	4.1E-02	9	AcCoAS, CG31075, CG32444, CG7430, CG9008, fbp, Pgi, Pglym78, Pgm1
dme00620	Glyoxylate and dicarboxylate metabolism	metabolism	carbohydrate	1.23E-02	4.4E-02	7	AcCoAS, CG7430, Gip, Irp-1B, Mdh1, Shmt, Spat
dme01212	Fatty acid metabolism	metabolism	lipid	2.20E-03	4.94E-02	9	ACC, ACOX1, Acsl, CG3902, CG9527, CPT2, Desat1, FASN1, Mtpalpha
dme04624	Toll and Imd signaling pathway	organismal systems	immune system	4.37E-03	5.1E-02	11	BomS1, BomS2, BomS3, Dif, DptA, DptB, Drs, Dso1, PGRP-LE, PGRP-SB1, spz

dme00620	Pyruvate metabolism	metabolism	carbohydrate	2.70E-02	6.6E-02	7	ACC, AcCoAS, CG31075, CG7430, Mdh1, Men, PCB
dme00052	Galactose metabolism	metabolism	carbohydrate	1.23E-02	8.5E-02	6	Akr1B, CG32444, CG7997, Gale, Pgm1, UGP
dme00071	Fatty acid degradation	metabolism	lipids	3.25E-03	8.6E-02	7	ACOX1, Acsl, CG31075, CG3902, CG9527, CPT2, Mtpalpha
dme00640	Propanoate metabolism	metabolism	carbohydrate	3.68E-02	8.9E-02	6	ACC, AcCoAS, ACOX1, CG7430, CG9527, Mtpalpha
dme00230	Purine metabolism	metabolism	nucleotide	2.81E-02	1.01E-01	10	CG11089, CG32549, cN-IIIB, dnc, Pde9, Pfas, Pgm1, Prat2, Prps, ras
dme04070	Phosphatidylinositol signaling	environmental information processing	signal transduction	4.73E-03	1.17E-01	14	Cam, Cds, CG34384, CG3530, CG42271, CG9391, EDTP, fwd, INPP5E, Itpr, norpA, Ocrl, Pkc53E, Plc21C
dme00562	Inositol phosphate metabolism	metabolism	carbohydrate	1.26E-02	1.29E-01	9	CG3530, CG42271, CG9391, EDTP, fwd, INPP5E, norpA, Ocrl, Plc21C
dme04080	Neuroactive ligand-receptor interaction	environmental information processing	signal transduction	2.68E-02	1.79E-01	6	5-HT1A, baz, CCKLR-17D1, GABA-B-R2, Lgr1, mAChR-B
dme00564	Glycerophospholipid metabolism	metabolism	lipid	3.32E-02	2.59E-01	9	Cds, CG2201, CG3009, CG34384, Gpo1, iPLA2-VIA, nes, Pect, Pld3
dme04150	mTOR signaling pathway	environmental information processing	signal transduction	2.97E-02	2.94E-01	10	Akt1, eIF4B, Jhl-21, Mo25, Pkc53E, Vha14-1, Vha44, Vha55, Vha68-1, Wnt6

Table S4 — Decathlon enriched KEGG pathways for all decathlon experiments. Class and sub-class correspond to KEGG pathway primary and secondary class terms. Observed *p*-value is the minimum adjusted *p*-value of the pathway enrichment across all behavioral metrics in the observed data (see method details). Average minimum *p*-value is the mean of such *p*-values across bootstrap replicates. Gene hits and predictive genes correspond to the number and list of genes associated with the KEGG pathway in the observed data.